Risque de crédit: modélisation et simulation numérique.

Ying Jiao

THÈSE DE DOCTORAT

présentée par

Jiao Ying

pour obtenir

le grade de : Docteur de l'École Polytechnique
Spécialité : Mathématiques Appliquées

Risque de crédit:
modélisation et simulation numérique

Thèse présentée le 11 décembre 2006 devant la commission d'examen :

Nicole	El Karoui	Directrice de thèse
Jean-Pierre	Fouque	Rapporteur
Emmanuel	Gobet	Examinateur
Monique	Jeanblanc	Rapportrice
David	Kurtz	Examinateur
Gilles	Pagès	Président du jury
Philip	Protter	Rapporteur
Nizar	Touzi	Examinateur

À ma mère

Remerciement

Je voudrais exprimer toute ma gratitude à Nicole El Karoui, ma directrice de thèse, qui m'a initiée dans le domaine de mathématiques financières. Elle m'a soutenue tout au long de la préparation de cette thèse. Son enthousiasme scientifique, sa vision sur les problèmes pratiques ainsi que leurs modélisations mathématiques, et son exigence sur la rédaction ont laissé sur moi beaucoup d'influences. Je tiens à la remercier profondément pour les temps qu'elle m'a consacrés et pour la confiance et les encouragements qu'elle m'a manifestés sans lesquelles cette thèse n'aurait pas pu sortir.

Je remercie sincèrement Jean-Pierre Fouque, Monique Jeanblanc et Philip Protter pour avoir accepté de rapporter cette thèse, malgré le delai relativement court, et pour avoir communiqué avec moi leurs remarques très précieuses.

Je tiens à remercier Emmanuel Gobet, David Kurtz, Gilles Pagès et Nizar Touzi pour avoir fait partie du jury et particulièrement Gilles Pagès pour l'avoir présidé.

Je remercie vivement Monique Jeanblanc pour l'intérêt qu'elle a porté sur mon travail, pour les questions qu'elle m'a posées et les références qu'elle m'a données, pour ses nombreux conseils sur mes manuscrits, et aussi pour ses encouragements constants.

Je remercie Emmanuel Gobet pour avoir toujours voulu partager avec moi ses connaissances et ses idées. Ses commentaires sont utiles et ses soutiens sont précieux pour cette thèse.

Certains résultats dans la deuxième partie de cette thèse ont été implémentés par David Kurtz et son équipe à Calyon. Je lui sais gré d'avoir communiqué leurs résultats empiriques avec moi.

La motivation pratique de la première partie de cette thèse est basée sur un problème posé, lors de mon stage de DEA effectué chez Crédit Agricole Indosuez, par mon directeur de stage, Yann Samuelidès. Je lui exprime ma gratitude. Je remercie également François Dezorme et Florent Serre pour leurs explications sur l'aspet pratique du marché pendant ce stage.

Noureddine El Karoui m'a envoyé des références sur l'approximation poissonnienne qui m'ont motivée à regarder des problèmes associés. Je lui sais gré.

Je voudrais remercier l'équipe de mathématiques financières, particulièrement les participants du groupe de travail au Centre de Mathématiques Appliquées de l'Ecole Polytechnique où cette thèse est effectuée.

Mes remerciements vont aussi à Jeanne, Nasséra, Sylvain, Aldjia et Véronique pour leurs accueils et aides chaleureux et efficaces.

Mille mercis à mes amis thésards et post-docs pour les amitiés et les soutiens, aussi pour les moments agréables que nous avons partagés.

Enfin, je remercie ma famille pour les soutiens morals permanents.

Résumé

Cette thèse est motivée par les problèmes induits par la corrélation des défauts dans les produits dérivés de crédit. La thèse contient deux parties. La première est consacrée à analyser théoriquement les défauts successifs. On propose une nouvelle approche, basée sur la densité des probabilités conditionnelles de survie, pour traiter ce qui se passe après le premier défaut en déterminant les compensateurs des défauts successifs et en calculant les espérances conditionnelles par rapport à la filtration du marché. Dans la deuxième partie, on présente une méthode d'approximation pour calculer les prix des CDOs en utilisant la méthode de Stein et la transformation de zéro biais. On obtient un terme correcteur explicite pour l'approximation gausienne et on estime la vitesse de convergence. Les tests numériques montrent l'efficacité de cette méthode par rapport aux méthodes classiques. On établit aussi des résultats similaires pour l'approximation poisonnienne en appuyant sur des variantes discrètes de la méthode. Enfin, pour les fonctions plus régulières, on propose des correcteurs d'ordres supérieurs.

Abstract

The thesis is motivated by the problems related to the defaults correlation in the portfolio credit derivatives. The thesis contains two parts. The first one is devoted to the analysis of successive defaults. We propose a new approach, based on the density of the conditional survival probabilities, to study the phenomena after the first default by determining the compensators of successive defaults and by calculating the conditional expectations with respect to the global filtration of the market. In the second part, we present an approximation method the evaluate the CDOs, using the Stein's method and the zero bias transformation. We obtain a correction term for the normal approximation and we estimate the convergence speed. Numerical tests show the efficiency of this method compared to the classical methods. We also establish similar results for Poisson approximation by adopting a discrete version of the method. Ar last for more regular functions, we proposer high-ordered corrections.

Contents

Notations

Part I

$(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$ is a filtered probability space which represents the market and the filtration $(\mathcal{G}_t)_{t\geq 0}$ represents the global market information, denoted by $\mathbb{G}$.

The single-credit case

- τ is a $\mathbb{G}$-stopping time.
- $\mathbb{D} = (\mathcal{D}_t)_{t\geq 0}$ is the filtration defined by $\mathcal{D}_t = \sigma(1\!\!1_{\{\tau\leq s\}}, s \leq t)$.
- $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is a sub-filtration of $\mathbb{G}$.
- Λ is the $\mathbb{G}$-compensator process of τ; $\Lambda^{\mathbb{F}}$ is the $\mathbb{F}$-predictable process which coincides with Λ on $\{\tau > t\}$.
- G is the $\mathbb{F}$-survival process defined by $G_t = \mathbb{P}(\tau > t|\mathcal{F}_t)$.
- G^θ is the process of the conditional survival probability defined by $G_t^\theta = \mathbb{P}(\tau > \theta|\mathcal{F}_t)$.
- $\mathcal{E}(X)$ is the Doléans-Dade exponential of the semi-martingale X.
- $\alpha_t(\theta)$ is the density of G_t^θ.
- q_t is the $\mathcal{F}_t$-measure defined by $q_t(f) = \mathbb{E}[f(\tau)|\mathcal{F}_t]$.

The two-credits case

- (τ_1, τ_2) is a family of two $\mathbb{G}$-stopping times.
- $\tau = \tau_1 \wedge \tau_2$, $\sigma = \tau_1 \vee \tau_2$.
- $\mathbb{D}^1$, $\mathbb{D}^2$, $\mathbb{D}^\tau$ and $\mathbb{D}^\sigma$ are natural filtrations associated to the processes associated with $\mathbb{G}$-stopping times τ_1, τ_2, τ and σ.
- $\mathbb{D} = \mathbb{D}^1 \vee \mathbb{D}^2$, $\mathbb{D}^{\tau,\sigma} = \mathbb{D}^\tau \vee \mathbb{D}^\sigma$.
- $\mathbb{F}$ is a sub-filtration of $\mathbb{G}$ such that $\mathbb{G} = \mathbb{D} \vee \mathbb{F}$.
- $\mathbb{G}^i = \mathbb{F} \vee \mathbb{D}^i$.
- $\mathbb{G}^\tau = \mathbb{D}^\tau \vee \mathbb{F}$, $\mathbb{G}^{\tau,\sigma} = \mathbb{D}^{\tau,\sigma} \vee \mathbb{F}$.
- Λ^i (resp. Λ^τ, Λ^σ) is the $\mathbb{G}$-compensator process of τ_i (resp. τ, σ);
- $H_t = \mathbb{P}(\sigma > t|\mathcal{G}_t^\tau)$, $H_t^\theta = \mathbb{P}(\sigma > \theta|\mathcal{G}_t^\tau)$.
- p_t is the density of the joint conditional probability $\mathbb{P}(\tau > u, \sigma > v|\mathcal{F}_t)$.
- α_t is the density of the joint conditional probability $\mathbb{P}(\tau_1 > t_1, \tau_2 > t_2|\mathcal{F}_t)$.

Part II

- L_T is the cumulative loss up to time T.
- Y is the common factor and Y_i is the individual factor.
- X^* represents a r.v. having the X-zero bias distribution where X is a zero-mean r.v..
- $W = X_1 + \cdots + X_n$ where X_i is a zero-mean r.v. $(i = 1, \cdots, n)$.
- Z represents a zero-mean normal r.v..
- $\Phi_\sigma(h) = \mathbb{E}[h(Z)]$ where $Z \sim N(0, \sigma^2)$.
- f_h and $f_{h,\sigma}$ represent the solution of the Stein's equation.
- $\widetilde{f}_h$ and $\widetilde{f}_{h,\sigma}$ represent the solution of the decentralized Stein's equation.
- $I_\alpha(x) = \mathbb{1}_{\{x \leq \alpha\}}$.
- $C_k(x) = (x - k)^+$.
- $\mathbb{N}^+$ represents the non-negative integer set.
- p_h and $p_{h,\sigma}$ represent the solution of the Stein's Poisson equation.
- $\widetilde{p}_h$ and $\widetilde{p}_{h,\sigma}$ represent the solution of the decentralized Stein's Poisson equation.
- $\delta(N, f, X, Y)$ represents the remaining term of the N^{th} order Taylor's formula in the expectation form.
- $\varepsilon(N, f, X, Y)$ represents the remaining term of the N^{th} order reversed Taylor's formula in the expectation form.

Introduction de la thèse

0.1 Introduction

Cette thèse est motivée par des problèmes liés à la modélisation du risque de crédit dans un univers qui concerne plusieurs entreprises. La grande question posée par le marché est la corrélation entre les faillites des diverses entreprises. C'est un sujet de grande importance en finance de marché car le risque de crédit est un risque systématique qui dépend du cycle économique; en particulier, les défauts des entreprises dans le même secteur ou dans la même région géographique sont fortement corrélés. On peut trouver dans la littérature sur le risque de crédit (Bruyère et al.[13], Duffie et Singleton [30] et Bluhm, Overbeck et Wagner[10], etc.), une étude général de problèmes associés à ce thème. Pour rendre possible une gestion plus flexible et efficace de ce type de risque, des instruments financiers basés sur un panier de crédits ont été proposés. Ils peuvent en gros être classés dans deux catégories : les "basket default swaps" et les "collateralized debt obligations (CDO)".
Récemment, des indices de portefeuilles de crédit synthétiques ont été introduits, permettant aux investisseurs d'avoir plus de liquidité et plus d'instruments pour se couvrir.

Différents types de problèmes mathématiques se posent pour étudier le risque induit par la corrélation de défauts. Deux d'entre eux sont traités dans cette thèse : le premier est l'analyse de ce qui se passe après le premier défaut pour un portefeuille comprenant plusieurs entreprises; le deuxième est de donner une méthode numérique rapide et robuste pour évaluer les CDOs.

Une grande famille de modèles de défaut repose sur une représentation de la loi conditionelle du temps de défaut. La modélisation du temps de défaut a été très étudiée, voir par example Bélanger, Shreve et Wong [8], Elliott, Jeanblanc et Yor [31], Jeanblanc et Rutkowski [55], Zhou [82], ainsi que la monographie de Bielecki et Rutkowski [9]. La littérature sur les multi-défauts, comportant en particulier Schönbucher et Schubert [74] et Lando [59] ne présente pas une modélisation dans un cadre général, ce problème étant plus complexe à étudier.

Quand on considère un portefeuille qui contient des actions émises par plusieurs entreprises susceptibles de faire faillite dans ce cadre de modélisation, une question

posée par les praticiens est la suivante : *puisqu'on observe dans le marché essentiellement le premier défaut, peut-on considérer que les défauts futurs vont garder une loi de la même famille, mais dont le paramètre dépend du temps d'observation.*
La première partie de cette thèse est motivée par cette question. On rappelle d'abord certains résultats obtenus dans l'approche de type intensité, en faisant une présentation adaptée aux extensions au cas multi-crédits; on introduit une condition un peu plus générale sur les filtrations concernées comme dans Jeulin et Yor [57] et Jacod [49] ou proposée dans Guo, Jarrow et Menn [44] and Guo et Zeng [46]. En particulier, on s'intéresse à la décomposition multiplicative de la surmartingale dite "processus de survie" qui nous permet d'obtenir des développements de type HJM, sous une probabilité bien choisie. Tous ces résultats permettent de caractériser les probabilités de survie avant le défaut. L'étude de ce qui se passe après le défaut est présenté dans Bielekie, Jeanblanc et Rutkowski [79]. Nous avons mené une étude exhaustive de cet aspect à partir de la loi $\mathcal{F}_t$-conditionnelle de survie restreinte à $[0, t]$. Cela permet de représenter les martingales de la tribu minimale engendrée par l'information a priori et le défaut avant t, soit $t \wedge \tau$.
Dans le chapitre suivant, nous revenons à la question des praticiens, en proposant un modèle simple dans un cadre d'intensités déterministes pour étudier le problème; on montre que la propriété proposée par les practiciens ne peut avoir lieu que sous des hypothèses très particulières. La distribution jointe et la copule associée sont très différentes de celles utilisées en général sur le marché. Nous montrerons les conséquences pratiques que cela induit.
Dans le cas où les intensités sont stochastiques, on ne peut pas généraliser facilement le point de vue du marché, qui correspond donc à une intuition erronée. Par contre, on peut montrer comment l'intensité du deuxième défaut va dépendre de celle du premier, et donner quelques propriétés caractéristiques. Même avec seulement deux noms, les calculs sont vite très compliqués. Pourtant, en adoptant le cadre général introduit dans le premier chapitre, en prenant comme filtration de l'information celle générée par l'information *a priori* et celle additionnelle due à l'observation du défaut, nous pouvons étudier le processus d'intensité du deuxième défaut. Les résultats de ce chapitre sont nouveaux. Une hypothèse importante permet de mener l'étude jusqu'au bout, c'est celle qui dit que la loi conditionnelle du défaut avant t admet une densité par rapport à la mesure de Lebesgue.

Concernant les produits de grande taille comme les CDOs, la recherche d'une méthode numérique efficace reste priviligiée. Sur le marché, les practiciens adoptent une approche simplifiée où les temps de défauts sont supposés être corrélés par un seul facteur commun, et conditionnellement à ce facteur, les défauts sont indépendants. Alors, la perte cumulative, qui est le terme clé pour évaluer une tranche de CDO, peut être calculée en deux étapes dans ce contexte : on calcule d'abord la perte conditionnelle qui peut s'écrire comme la somme des variables aléatoires indépendantes, et puis en prenant l'espérance de fonctions de la perte conditionelle, on obtient le prix de ce produit financier. Dans le cadre du modèle de facteur, l'approximation normale, qui est un résultat direct du théorème de la limite centrale, est proposée par Vasicek [80] et Shelton [75]. L'estimation de la vitesse de convergence est un problème classique dans l'étude du théorème de la limite centrale, voir par exemple l'inégalité de Berry-Esseen. Mais la difficulté majeure, dans le monde de crédit, est que l'on rencontre souvent des probabilités très petites pour lesquelles l'approximation normale ne peut plus être robuste. Plusieurs travaux récents de Antonov, Mechkov et Misirpashaev [2], Dembo, Deuschel et Duffie [24], Glasserman, Kang et Shahabuddin [36] sont consacrés à la question d'améliorer la qualité de l'approximation en utilisant des différentes méthodes (les grandes déviations, la méthode du col, etc.). Des tests numériques ont montré une amélioration par rapport à l'approximation gaussienne classique, pourtant, les estimations des erreurs de ces approximations ne sont pas discutées. Par ailleurs, elles sont parfois assez couteuses en temps de calcul.

Dans cette thèse, on propose une autre approche pour traiter ce problème en utilisant la méthode de Stein, qui est une méthode très puissante pour étudier la différences des espérances d'une même fonction par rapport à deux lois différentes, notamment quand l'une des lois est normale ou poissonnienne. En combinant la méthode de Stein avec la technique de zéro-biais transformation proposée par Goldstein et Reinert [39], on obtient, en faisant un développement autour de la perte totale, un terme de correction à l'approximation normale classique de l'espérance d'une fonction d'une somme directe de variables aléatoires indépendantes. En exprimant l'erreur de l'approximation gaussienne d'une fonction régulière de la perte comme une différence entre les espérances d'une même fonction auxilliaire pour la distribution de la somme de variables aléatoires et pour sa transformation de zéro-biais, on trouve un terme correcteur à l'approximation gaussienne et on obtient l'ordre d'erreur corrigée (d'ordre $O(1/n)$ pour le cas homogène). Pour certaines fonctions moins régulières, comme la fonction $(x-k)^+$ qui est très importante en finance, la démonstration de ce résultat est plus délicate, et demande d'établir des inéqualités de concentration ainsi qu'une technique d'espérance conditionelle. Cette correction est aussi efficace quand les probabilités sont très petites que pour le cas symétrique. De plus, elle est valable pour le cas où les variables ne sont pas nécessairement de type Bernoulli et pas identiquement distribuées. D'autre part, ce terme de correction peut s'écrire sous la forme de l'espérance du produit de la fonction considérée et d'un polynôme, pour la loi gaussienne. Cela rend le calcul du correcteur explicite et rapide, surtout lorsqu'on considère l'approximation de l'espérance d'une fonction de la somme de variables aléatoires indépendantes de lois éventuellement différentes. Des tests numériques sont effectués et on constate une amélioration significative de la qualité de l'approximation. Par rapport à la méthode classique de Monte Carlo, le temps de calcul est réduit substantiellement en gardant une très bonne précision. Les résultats sont comparables à ceux obtenus par la méthode du col, mais plus facile à implémenter. On analyse enfin l'impact du paramètre de corrélation sur l'effet de la correction.

Sur le plan théorique, on propose une formule du développement asymptotique pour l'espérance d'une fonction plus régulière de la somme de variables aléatoires indépendantes. Les termes successifs dans le développement consistent en les espérances des fonctions sous la distribution gaussienne ou poissonnienne, ainsi que les moments des variables. Il existe une littérature vaste sur ce thème et les méthodes utilisées sont variées: des fonctions caractéristiques dans Hipp [48], Götze et Hipp [42], la méthode de Stein combinant avec l'expansion de Edgeworth dans Barbour [3], Barbour [4], la méthode de Lindeberg dans Borisov et Ruzankin [11]. En particulier, Barbour a utilisé une approche basée sur la méthode de Stein pour traiter le cas normal et le cas poissonnien par des techniques similaires dans [3] et [4]. On emploie dans cette thèse la transformation de zéro-biais et on estime l'erreur d'approximation en faisant référence au cas du premier ordre. Plus précisément, en observant l'expansion du premier ordre qui demande l'existence de la dérivée d'ordre deux de la fonction, et qui s'applique éventuellement sur la fonction call, on déduit la relation entre la régularité de la fonction considérée et l'ordre effectif (c'est-à-dire, l'ordre jusqu'auquel on peut améliorer le grandeur de la vitesse de convergence) des développements asymptotiques et on donne les conditions nécessaires sur la fonction pour assurer l'existence du développement jusqu'à l'ordre N, où N est un entier positif. Toujours dans le cadre de la méthode de Stein, on établit aussi un développement asymptotique de l'approximation poissonnienne par des techniques analogues. Concernant l'application financière, la correction de l'ordre supérieur permet encore d'améliorer l'approximation normale, ce qui est clairement montré par des tests numériques.

0.2 Structure de la thèse et résultats principaux

Sur le premier temps de défaut et après

Le premier chapitre commence par une revue des résultats principaux dans l'approche d'intensité, d'un point de vue plus général et mieux adapté aux extensions dans le cas multi-dimensionnel. Dans le premier paragraphe, on rappelle des notions et des propriétés basiques des processus prévisibles pour étudier, dans le paragraphe suivant, le $\mathbb{G}$-compensateur d'un $\mathbb{G}$-temps d'arrêt τ, qui est en fait la projection duale prévisible du processus $(\mathbb{1}_{\{\tau\leq t\}}, t \geq 0)$. On s'intéresse au calcul du compensateur et on étudie deux exemples simples mais importants: un dans le cas déterministe et l'autre avec l'hypothèse (H), qui est une hypothèse souvent supposée dans la modélisation de crédits et qui est équivalente à l'indépendance conditionelle de $\mathcal{G}_t$ et $\mathcal{F}_\infty$ sachant $\mathcal{F}_t$.

Dans le troisième paragraphe, on propose un cadre général sous lequel on travaillera ensuite. L'hypothèse importante est celle proposée sur les filtrations $\mathbb{G}$ et $\mathbb{F}$: pour tout

$t \geq 0$ et tout $U \in \mathcal{G}_t$, il existe $V \in \mathcal{F}_t$, tel que $U \cap \{\tau > t\} = V \cap \{\tau > t\}$ (cette hypothèse a été utilisée par Guo, Jarrow et Menn [44]). Cette condition permet de traiter un seul temps de défaut et le premier temps de défaut d'un portefeuille de la même manière sans ajouter la moindre difficulté. On rappelle ensuite un résultat classique, à savoir qu'il existe un processus $\mathbb{F}$-prévisible $\Lambda^{\mathbb{F}}$ qui coïncide avant τ avec le $\mathbb{G}$-compensateur de τ et que $\Lambda^{\mathbb{F}}$ est calculable à partir d'une surmartingale $G = (\mathbb{P}(\tau > t|\mathcal{F}_t), t \geq 0)$, dit "processus de survie", qui joue un rôle essentiel dans la suite. De plus, par la même méthode, on montre que le processus $\Lambda^{\sigma,\mathbb{F}}$, qui coïncide avant τ avec le $\mathbb{G}$-compensateur d'un autre temps d'arrêt σ est donné par l'équation $dB_t^{\sigma,\mathbb{F}} = G_{t-} d\Lambda_t^{\sigma,\mathbb{F}}$, où $B^{\sigma,\mathbb{F}}$ est le compensateur du processus $V^\sigma = (\mathbb{P}(\sigma > \tau \wedge t|\mathcal{F}_t), t \geq 0)$.
En tant qu'une $\mathbb{F}$-surmartingale, G admet une décomposition multiplicative

$$G = \mathcal{E}(\widetilde{\Gamma}) = \mathcal{E}(\widetilde{M}^{\Gamma})\mathcal{E}(-\Lambda^{\mathbb{F}}),$$

qui nous permet d'introduire un changement de probabilité par rapport auquel on peut généraliser les développements de type HJM, faits d'habitude sous l'hypothèse (H); c'est l'objet du quatrième paragraphe. On traite le problème d'une façon similaire à celle des taux d'intérêt et on compare avec les résultats dans Schönbucher [72] par un changement de probabilité. Enfin, on conclut qu'il suffit de connaître le processus G pour déduire les probabilités conditionnelles de survie $\mathbb{P}(\tau > T|\mathcal{G}_t)$.

Pour traiter le cas après le défaut, on étudie dans le cinquième paragraphe les probabilités de survie conditionnelle à t, restreinte à $[0, t]$ pour en déduire les espérances conditionelles en sachant que le défaut a eu lieu. On développe dans les trois sous-paragraphes respectivement le cas avec l'hypothèse (H), le cas plus général sans l'hypothèse (H) mais où $\mathbb{P}(\tau > \theta|\mathcal{F}_t)$ avec $\theta \geq 0$ admet une densité, et puis enfin le cas général. Dans le cas avec densité, on montre que les espérances conditionelles peuvent être calculées explicitement par

$$\mathbb{E}[Y(T,\tau)|\mathcal{G}_t]1\!\!1_{\{\tau>t\}} = \frac{\mathbb{E}\big[\int_t^\infty Y(T,u)\alpha_T(u)du|\mathcal{F}_t\big]}{\int_t^\infty \alpha_t(u)du} 1\!\!1_{\{\tau>t\}}$$

et

$$\mathbb{E}[Y(T,\tau)|\mathcal{G}_t]1\!\!1_{\{\tau\leq t\}} = \mathbb{E}\Big[Y(T,s)\frac{\alpha_T(s)}{\alpha_t(s)}|\mathcal{F}_t\Big]\Big|_{s=\tau} 1\!\!1_{\{\tau\leq t\}}.$$

Dans ce cas-là, le processus compensateur du temps de défaut τ est donné par la formule

$$d\Lambda_t = 1\!\!1_{]0,\tau]}(t)\frac{\alpha_t(t)}{G_t}dt = 1\!\!1_{]0,\tau]}(t)\frac{\alpha_t(t)}{\int_t^\infty \alpha_t(u)du}dt.$$

Dans le cas général, on introduit la notion de $\mathcal{F}_t$-mesure $q_t(f) = \mathbb{E}[f(\tau)|\mathcal{F}_t]$ et on montre que l'espérance conditionelle peut être calculée par une dérivée au sens de Radon-Nikodym.

Le deuxième chapitre est consacré à l'étude de plusieurs temps de défauts. Le premier paragraphe a pour objectif de répondre à la question posée par les praticiens du marché mentionnée précédemment. On développe un modèle simple déterministe de deux temps de défaut, basé sur l'hypothèse suivante : les probabilités de survie individuelles suivent des lois de type exponentiel avant le premier défaut où les paramètres dépendent de l'observation du marché, soit $\mathbb{P}(\tau_i > T|\tau > t) = e^{-\mu^i(t)\cdot(T-t)}$, $(i = 1, 2)$. On montre alors que la probabilité jointe est déterminée complètement par les fonctions μ^i:

$$\mathbb{P}(\tau_1 > t_1, \tau_2 > t_2) = \exp\Big(- \int_0^{t_1} \mu^1(s \wedge t_2)ds - \int_0^{t_2} \mu^2(s \wedge t_1)ds\Big).$$

Ce résultat contredit l'hypothèse standard que la fonction de copule ne dépend pas de distributions marginales. Par ailleurs, cette probabilité jointe correspond à une fonction de copule spécifique $\widetilde{C}(u, v) = uv\rho\Big(\frac{\ln u}{\mu^1(0)}, \frac{\ln v}{\mu^2(0)}\Big)$ si $u, v > 0$ et $\widetilde{C}(u, v) = 0$ si $u = 0$ ou $v = 0$, qui ne ressemble pas à une forme utilisée habituellement sur le marché. On déduit ensuite respectivement la loi conditionelle du crédit qui survit et du deuxième temps de défaut et on observe que la propriété proposée par le marché n'est pas satisfaite dans ce cas et que les calculs deviennent vite compliqués. Donc on montre par ce modèle simple que l'intuition du marché est fausse et la corrélation de défaut demande une étude rigoureuse déjà dans le cas de deux temps de défaut.

Dans le deuxième paragraphe, on traite deux temps de défaut dans le cadre général. Duffie [27] montre que, sous certaine condition, le compensateur du premier temps de défaut est égal à la somme des compensateurs arrêtés à τ de chaque temps de défaut. Pour le deuxième temps de défaut σ, en appuyant sur les résultats obtenus dans le premier chapitre, on obtient son processus de compensateur :

$$d\Lambda_t^\sigma = 1\!\!1_{[\tau,\sigma]}(t)\frac{p_t(\tau, t)}{\int_t^\infty p_t(\tau, v)dv}dt,$$

où $p_t(u, v)$ est la densité de la probabilité conditionnelle jointe $\mathbb{P}(\tau > u, \sigma > v|\mathcal{F}_t)$ qui joue un rôle essentiel. Cette approche peut être étendue facilement en cas général pour les défauts successifs. Dans le dernier sous-paragraphe, on étend la méthode pour calculer les espérances conditionelles dans le chapitre précédent au cas de deux et de plusieurs crédits respectivement. En introduisant une famille de $\mathcal{F}_t$-mesures associées à chaque scénario de défaut, on calcule les espérances conditionelles par rapport à $\mathcal{G}_t$ comme des dérivées de type Radon-Nikodym. Cette méthode propose une possibilité de traiter le multi-crédits en ramenant aux calculs sur les espérances conditionelles par rapport aux tribus dans la filtration $\mathbb{F}$ que l'on pourrait éventuellement supposer être engendrée par un mouvement Brownien.

L'approximation de la perte cumulative

Le troisième chapitre de cette thèse traite le problème de l'approximation de la perte cumulative, qui est naturellement motivé par le besoin d'évaluer des produits dérivés de grande taille. Le premier paragraphe commence par une brève introduction au problème. On présente d'abord le modèle à facteur et puis l'étude des sommes de variables aléatoires indépendantes qui est l'objet principal à étudier dans ce contexte. Ensuite, on fait un rappel de la littérature sur la méthode de Stein et sur la transformation de zéro-biais, qui sont les outils que nous utiliserons pour traiter le problème de l'approximation.

Le deuxième paragraphe du chapitre est consacré à la transformation de zéro-biais et ses propriétés fondamentales. On rappelle d'abord les définitions et quelques résultats dûs à Goldstein et Reinert [39], puis on présente un exemple important pour la suite : les variables aléatoires d'espérance nulle qui suivent la distribution de Bernoulli asymétrique $\mathcal{B}(q,-p)$. Ensuite, on montre que si X est une variable aléatoire et si X^* est une autre variable aléatoire indépendante de X suivant sa loi de zéro-biais, alors on peut calculer explicitement l'espérance d'une fonction g de la différence entre X et X^* lorsque la fonction g est paire et est localement intégrable (cf. Proposition 3.2.6):

$$\mathbb{E}\big[g(X^*-X)\big] = \frac{1}{2\sigma^2}\mathbb{E}\big[X^s G(X^s)\big].$$

Ce résultat s'applique donc aux fonctions $\mathbb{E}[|X^*-X|]$ pour mesurer la distance entre X et X^* avec la norme L^1. Dans le cas où W est la somme de variables aléatoires indépendantes $X_1,\cdots,X_n$, si pour tout entier $1\leq i\leq n$ on note X_i^* une variable aléatoire independante des $X_1,\cdots,X_n$ suivant la loi de zéro-biais de X_i, alors la variable aléatoire W^* définie par $W^* = W^{(I)} + X_I^*$ suit la loi de zéro-biais de W, où I est un indice, indépendant des X_i et des X_i^*, à valeur dans $\{1,\cdots,n\}$ et dont la loi est déterminée par les variances des X_i (cf. [39]). Le point important est que dans ce cas on n'a plus l'indépendance entre W et W^*. Cette difficulté est levée en montrant que les covariances des quantités à estimer ne dépendent que de la variance de l'espérance contionnelle de $(X_I - X_I^*)$ sachant $(\overrightarrow{X},\overrightarrow{X}^*)$ qui est d'un ordre inférieure à celle de $(X_I - X_I^*)$. On donne une estimation importante dans la proposition 3.2.16 de la covariance de deux variables aléatoires dont l'une est une fonction de W et l'autre est une fonction de X_I et de X_I^*:

$$\begin{aligned}&\Big|\mathbb{E}[f(W)g(X_I,X_I^*)] - \mathbb{E}[f(W)]\mathbb{E}[g(X_I,X_I^*)]\Big|\\ &\leq \frac{1}{\sigma_W^2}\mathrm{Var}[f(W)]^{\frac{1}{2}}\Big(\sum_{i=1}^n \sigma_i^4 \mathrm{Var}[g(X_i,X_i^*)]\Big)^{\frac{1}{2}}.\end{aligned}$$

On remarque que l'utilisation directe de l'inégalité de Cauchy-Schwarz ne permet pas de trouver l'estimation suffisamment précise pour notre application. Il faut prendre l'espérance conditionnelle avant d'appliquer l'inégalité de Cauchy-Schwarz.

Le troisième paragraphe consiste à introduire l'équation de Stein associée à une fonction et à étudier les solutions de cette équation. Si h est une fonction régulière, l'équation de Stein associée à h est par définition

$$xf_h(x) - \sigma_W^2 f_h'(x) = h(x) - \Phi_{\sigma_W}(h).$$

Par abus de language, on utilisera $f_h(x)$ pour désigne la solution de l'équation de Stein qui croît le moins vite à l'infini.

En combinant ceci avec la définition de transformation de zéro-biais, l'équation de Stein donne

$$\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h) = \mathbb{E}[Wf_h(W) - \sigma_W^2 f_h'(W)] = \sigma_W^2 \mathbb{E}[f_h'(W^*) - f_h'(W)].$$

Ceci ramène l'étude sur l'erreur de l'estimation gausienne de $\mathbb{E}[h(W)]$ à une étude sur la différence des espérances de la même fonction f_h' en deux variables aléatoires différentes (mais très similaires).

L'estimation de l'erreur de l'approximation gaussienne repose donc sur l'estimation de la distance entre les variables aléatoires W et W^* qui a été discutée dans le paragraphe précédent et sur le contrôle des croissances des dérivées de la fonction f_h.

On développe deux méthodes dont l'une est inspirée par Barbour [3] pour estimer les croissances des dérivées de la fonction f_h (comme par exemple la norme sup des fonctions $|f_h'|$, $|xf_h'|$, $|xf_h''|$, etc.) pour une certaine fonction h. En particulier, on étudie le cas où h est la fonction indicatrice ou la fonction call.

Ayant obtenu les estimations nécessaires dans les deux paragraphes précédents, on démontre dans le paragraphe 3.4 les résultats principaux du chapitre. Tout d'abord, les estimations du premier ordre sont données pour des fonctions avec différentes conditions de régularité et diverses vitesses de croissance: le lemme 3.4.1 concerne des fonctions h dont les dérivées sont bornées et le lemme 3.4.2 traite des fonctions h dont les dérivées sont à croissance linéaire. La proposition 3.4.6 s'intéresse au cas où h est une fonction indicatrice, la preuve est basée sur l'inéqualité de concentration inspirée par Chen and Shao [17]. Ensuite, le théorème principal (Théorème 3.4.8) est établi pour le cas où la fonction h est lipschitzienne et la dérivée d'ordre trois de f_h existe et est bornée — on propose une amélioration de l'approximation gaussienne en ajoutant un terme correcteur

$$C_h = \frac{1}{\sigma_W^2}\mathbb{E}[X_I^*]\Phi_{\sigma_W}\left(\Big(\frac{x^2}{3\sigma_W^2} - 1\Big)xh(x)\right).$$

L'estimation de l'erreur corrigée est donnée par

$$\begin{aligned}&\Big|\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h) - C_h\Big|\\ &\leq \|f_h^{(3)}\|\left(\frac{1}{12}\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^4\big] + \frac{1}{4\sigma_W^2}\Big|\sum_{i=1}^n \mathbb{E}[X_i^3]\Big|\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big] + \frac{1}{\sigma_W}\sqrt{\sum_{i=1}^n \sigma_i^6}\right).\end{aligned}$$

Les variables aléatoires X_i sont indépendantes, mais ne sont pas nécessairement identiquement distribuées. Dans le cas où tous les X_i suivent la loi de Bernoulli asymétrique du même paramètre, la borne de l'erreur est de l'ordre $O(1/n)$. Cette correction est intéressante notamment pour des petites (ou grandes) probabilités car elle permet de trouver la même vitesse de convergence que dans le cas symétrique (lorsque tous les X_i suivent la loi Bernoulli symétrique) où la correction est automatiquement nulle.

Ce théorème, pourtant, ne peut pas s'appliquer à la fonction qui nous intéresse: la fonction call, car elle ne possède pas de dérivée d'ordre deux, et par conséquent la solution de l'équation de Stein associée ne possède pas de dérivée d'ordre trois. Ce problème est résolu dans la Proposition 3.4.15 en utilisant l'inégalité de concentration. On montre que C_h donné au-dessus reste valable dans le cas de la fonction call, et l'erreur de l'approximation corrigée est de bon ordre:

$$
\begin{aligned}
&\left|\mathbb{E}[(W-k)^+] - \Phi_{\sigma_W}((x-k)^+) - C_{(x-k)^+}\right| \\
&\leq \frac{1}{\sigma_W^2} \sum_{i=1}^n \Big(\frac{\mathbb{E}[|X_i^s|^4]}{3} + \sigma_i \mathbb{E}[|X_i^s|^3] \Big) \\
&+ \frac{1}{4\sigma_W^2} \sum_{i=1}^n \mathbb{E}[|X_i^s|^3] \Big(\frac{2\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]}{\sigma_W^3} + \frac{\Big(\sum_{i=1}^n \sigma_i \mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}}{\sqrt{2}\sigma_W^2} \Big) \\
&+ \mathrm{Var}[f''_{C_k}(W)]^{\frac{1}{2}} \Big(\sum_{i=1}^n \sigma_i^6 \Big)^{\frac{1}{2}} + \frac{1}{4\sigma_W^2} \sum_{i=1}^n \mathbb{E}[|X_i^s|^3] \Big(B(W,k) + \frac{c}{2\sigma_W^2} \sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big] \Big).
\end{aligned}
$$

La suite de ce paragraphe est consacrée aux tests numériques. Des comparaisons sont faites sur des portefeuilles homogènes et non-homogènes où les variables aléatoires suivent des lois de Bernoulli identiques ou non-identiques respectivement. Des résultats numériques montrent une amélioration substantielle de la qualité de l'approximation gaussienne.

On compare numériquement notre méthode à la méthode du col. Notre méthode conduit à une précision meilleure que la correction du premier ordre de la méthode de col. Bien que l'estimation gaussienne corrigée au deuxième ordre de la méthode de col soit plus précise que la nôtre, il faut noter que notre formule est beaucoup plus facile à calculer, surtout dans le cas non-homogène.

Dans le dernier paragraphe du chapitre, on étudie l'impact du facteur commun. Des tests numériques permettent de conclure que la correction est efficace lorsque la corrélation est peu importante. Dans le cas où la corrélation est forte, après avoir intégré par rapport au facteur, l'approximation normale reste robuste par effet de compensation.

Développements asymptotiques pour la distribution gaussienne et poissonnienne

Le quatrième chapitre est consacré à l'étude du développement asymptotique de $\mathbb{E}[h(W)]$, où W est la somme de variables aléatoires indépendantes. C'est une extension des résultats obtenus dans le troisième chapitre. Rappelons que l'on peut améliorer l'approximation gaussienne de $\mathbb{E}[h(W)]$ en ajoutant un terme de correction lorsque h possède certaines propriétés de régularité. Il est donc naturel d'espérer que lorsque h a des dérivées d'ordres supérieurs, on peut obtenir des termes de correction d'ordres correspondants. En développant des techniques dans le cadre de la méthode de Stein et de la transformation de zéro-biais, on propose une nouvelle approche pour traiter ce problème classique. Plus précisément, on peut résumer notre résultats en les trois points suivants:

1) On propose une "formule de Taylor" spéciale ayant deux versions — continue et discrète — qui permet d'obtenir des résultats similaires dans le cas normal et le cas poissonnien, respectivement;

2) Dans le cas normal, on donne les conditions nécéssaires sur la régularité et la croissance à l'infini de la fonction h qui permettent d'obtenir le développement d'ordre supérieur, en s'appuyant sur la "formule de Taylor" que l'on introduit. On discute l'estimation de l'erreur après correction et on en déduit que la convergence de l'approximation corrigée est de bon ordre.

3) Dans le cas poissonnien, on étend la notion de la transformation de zéro-biais aux variables aléatoires prenant valeurs dans $\mathbb{N}^+$. En utilisant la version discrète de la "formule de Taylor" on obtient le développement complet qui est similaire à celui obtenu dans le cas gaussien.

Le chapitre commence par un bref rappel sur les résultats de la littérature. Dans le paragraphe 4.2, on propose d'abord la première méthode utilisant la version classique de la formule de Taylor. Nous observons que dans le développement de Taylor de $\mathbb{E}[f_h(W)]$ ou de $\mathbb{E}[f_h(W^*)]$, les termes $\mathbb{E}[f_h^{(k)}(W^{(i)})]$ apparaissent naturellement pusique $W^{(i)}$ est indépendant de X_i ou de X_i^*. Il est donc possible de remplacer ces termes par leurs approximations gaussiennes, corrigées par le correcteur du premier ordre. On obtient ainsi un terme de correction d'ordre 2. En itérant le procédé ci-dessus, on obtient par récurrence une formule de développement à l'ordre quelconque, si la régularité de h est suffisamment forte. Pourtant, dans chaque étape de récurrence, on élimine une composante dans W, par exemple, au lieu d'approcher l'espérance d'une fonction de W, on estime $\mathbb{E}[f_h^{(k)}(W^{(i)})]$ dans la première étape. Par conséquent, dans la formule de développement, il apparait des sommes partielles de variables qui compliquent les calculs.

Une façon de se débarraser de la difficulté ci-dessus est de remplacer $\mathbb{E}[f_h^{(k)}(W^{(i)})]$ par une formule où il n'intervient que les espérances de fonctions de W. Comme

$W^{(i)} = W - X_i$, il est naturel de penser à réutiliser la formule de Taylor classique sur W. Auparavant, quand on applique l'espérance sur cette formule de Taylor, on ne peut pas obtenir la forme souhaitée car W et X_i ne sont plus indépendantes. Il est donc nécessaire de proposer une nouvelle formule pour approximer l'espérance d'une fonction de $W^{(i)}$ à l'ordre quelconque, telle que, dans la formule, il apparaisse seulement les espérances de fonctions de W.

La formule ci-dessous est la formule clé du dernier chapitre:

$$\mathbb{E}[f(X)] = \mathbb{E}[f(X+Y)] + \sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} \mathbb{E}[f^{(|\mathbf{J}|)}(X+Y)]\Big(\prod_{l=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}\Big) + \varepsilon(N,f,X,Y),$$

où $|\mathbf{J}| = j_1 + \cdots + j_d$ pour tout $\mathbf{J} = (j_l) \in \mathbb{N}_*^d$. Cette formule est une variante rétrograde de la formule de Taylor. Le terme d'erreur ε peut être calculé à partir des termes d'erreur de la formule de Taylor classique, appliquée sur $f^{(k)}(X+Y)$ (le développement est en X). Plus précisément, si on désigne par $\delta(N,f,X,Y)$ le nombre défini par l'égalité

$$\mathbb{E}[f(X+Y)] = \sum_{k=0}^{N}\frac{\mathbb{E}[Y^k]}{k!}\mathbb{E}[f^{(k)}(X)] + \delta(N,f,X,Y),$$

alors on a la relation

$$\varepsilon(N,f,X,Y) = -\sum_{d\geq 0}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} \delta(N-|\mathbf{J}|, f^{(|\mathbf{J}|)}, X, Y)\prod_{l=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}.$$

Intuitivement, sous des conditions convenables sur f, X et Y, $\varepsilon(N,f,X,Y)$ doit avoir le même ordre que $\delta(N,f,X,Y)$.

En prenant $f = f_h^{(k)}$, $X = W^{(i)}$ et $Y = X_i$, cette formule clé donne une estimation de $\mathbb{E}[f_h^{(k)}]$ où apparaissent des expressions de la forme $\mathbb{E}[f_h^{(l)}(W)]$ et $\mathbb{E}[Y^m]$. La méthode de récurrence présentée ci-dessus complétée par ce procédé technique donne l'estimation suivante de $\mathbb{E}[h(W)]$, qui est de l'ordre N quelconque (cf. Théorème 4.2.5):

Si N est un entier positif, on peut écrire $\mathbb{E}[h(W)]$ sous la forme $\mathbb{E}[h(W)] = C(N,h) + e(N,h)$, où

1) $C(0,h) = \Phi_{\sigma_W}(h)$ et $e(0,h) = \mathbb{E}[h(W)] - \Phi_{\sigma_W}(h)$;

2) et par récurrence, pour tout $N \geq 1$,

$$C(N,h) = \Phi_{\sigma_W}(h) + \sum_{i=1}^{n}\sigma_i^2 \sum_{d\geq 1}(-1)^{d-1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} C(N-|\mathbf{J}|, f_h^{(|\mathbf{J}|+1)})$$
$$\left(\prod_{l=1}^{d-1}\frac{\mathbb{E}[X_i^{j_l}]}{j_l!}\right)\left(\frac{\mathbb{E}[(X_i^*)^{j_d}]}{j_d!} - \frac{\mathbb{E}[X_i^{j_d}]}{j_d!}\right),$$

$$
\begin{aligned}
&e(N,h)\\
&=\sum_{i=1}^{n}\sigma_i^2\sum_{d\geq 1}(-1)^{d-1}\sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} e\big(N-|\mathbf{J}|,f_h^{(|\mathbf{J}|+1)}\big)\left(\prod_{l=1}^{d-1}\frac{\mathbb{E}[X_i^{j_l}]}{j_l!}\right)\left(\frac{\mathbb{E}[(X_i^*)^{j_d}]}{j_d!}-\frac{\mathbb{E}[X_i^{j_d}]}{j_d!}\right)\\
&+\sum_{i=1}^{n}\sigma_i^2\sum_{k=0}^{N}\frac{\mathbb{E}[(X_i^*)^k]}{k!}\varepsilon\big(N-k,f_h^{(k+1)},W^{(i)},X_i\big)+\sum_{i=1}^{n}\sigma_i^2\delta\big(N,f_h',W^{(i)},X_i^*\big)
\end{aligned}
$$

En particulier, si $N=1$, on retrouve le terme correcteur que l'on obtient dans le premier chapitre.

Des tests numériques montrent que l'approximation corrigée à l'ordre deux est de même précision que celle de la méthode de col, lorsque W est la somme de variables aléatoires indépendantes suivant la même loi de Bernoulli asymétrique et lorsque h est la fonction call.

L'erreur de l'approximation est estimée dans le sous-paragraphe 4.2.3. Si on développe la formule de récurrence qui définit $C(N,h)$, les dŕivées d'ordres supérieurs de la solution de l'équation de Stein associée vont apparaître. Par conséquent, pour que $C(N,h)$ soit bien défini, il faut que la fonction h soit dans un espace de fonctions sur lequel toute composée de longueur convenable d'opérateurs de la forme $\varphi\mapsto f_\varphi^{(l)}$ est bien définie. En particulier, h doit être suffisamment régulière, et la croissance de h à l'infini ne doit pas être trop grande. La première partie du sous-paragraphe consiste à définir les espaces de fonctions avec lesquels on va travailler et à discuter des propriétés agissant sur ces espaces de fonctions. C'est une préparation à l'estimation du terme d'erreur qui se trouve à la fin du sous-paragraphe.

La difficulté majeure de l'estimation du terme d'erreur est déjà apparue dans le chapitre précédent. Ici la vitesse de croissance de $\widetilde{f}_h^{(N+2)}$ joue un rôle crucial pour le développement de l'ordre N (rappelons que dans le chapitre précédent avec $N=1$, le comportement de la fonction $\widetilde{f}_h^{(3)}$ était essentiel pour l'estimation du terme d'erreur). L'un des objectifs de la deuxième partie du sous-paragraphe 4.2.3 est de proposer des conditions sur h avec lesquelles on peut obtenir les propriétés désirées de $\widetilde{f}_h^{(N+2)}$.

Enfin, l'estimation de l'erreur est donnée par une formule de récurrence (cf. Proposition 4.2.23). Avec ceci on montre aisément que (cf. Proposition 4.2.25), dans le cas de Bernoulli asymétrique où les variables aléatoires sont identiquement distribuées, l'ordre de l'erreur est donné par :

$$
e(N,h)\sim O\left(\left(\frac{1}{\sqrt{n}}\right)^{N+1}\right).
$$

Dans la section 4.3, on applique la méthode précédente sur la loi de Poisson en proposant une variante discrète de la "formule de Taylor" rétrograde (4.55):

$$\mathbb{E}[f(X)] = \mathbb{E}[f(X+Y)] + \sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}|} f(X+Y)] \prod_{l=1}^{d} \mathbb{E}\left[\binom{Y}{j_l}\right]$$
$$+\, \varepsilon(N, f, X, Y)$$

où $|\mathbf{J}| = j_1 + \cdots + j_d$ pour tout $\mathbf{J} = (j_l) \in \mathbb{N}_*^d$.

La forme de la formule de développement ainsi que la démonstration sont très similaires à celles dans le cas normal. Si N est un entier positif, alors on peut écrire $\mathbb{E}[h(W)]$ sous la forme $\mathbb{E}[h(W)] = C(N,h) + e(N,h)$, où

1) $C(0,h) = \mathcal{P}_{\lambda_W}(h)$ et $e(0,h) = \mathbb{E}[h(W)] - \mathcal{P}_{\lambda_W}(h)$,

2) et par récurrence, pour tout $N \geq 1$,

$$C(N,h) = \mathcal{P}_{\lambda_W}(h) + \sum_{i=1}^{n} \lambda_i \sum_{d\geq 1}(-1)^{d-1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} C(N-|\mathbf{J}|, \Delta^{|\mathbf{J}|} p_h(x+1))$$
$$\left(\prod_{l=1}^{d-1} \mathbb{E}\left[\binom{X_i}{j_l}\right]\right) \mathbb{E}\left[\binom{X_i^*}{j_d} - \binom{X_i}{j_d}\right],$$

$$\begin{aligned} &e(N,h) \\ &= \sum_{i=1}^{n} \lambda_i \sum_{d\geq 1}(-1)^{d-1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} e(N-|\mathbf{J}|, \Delta^{|\mathbf{J}|} p_h(x+1)) \prod_{l=1}^{d-1} \mathbb{E}\left[\binom{X_i}{j_l}\right] \mathbb{E}\left[\binom{X_i^*}{j_d} - \binom{X_i}{j_d}\right] \\ &+ \sum_{i=1}^{n} \lambda_i \sum_{k=0}^{N} \mathbb{E}\left[\binom{X_i^*}{k}\right] \varepsilon(N-k, \Delta^k p_h(x+1), W^{(i)}, X_i) + \sum_{i=1}^{n} \lambda_i \delta(N, p_h(x+1), W^{(i)}, X_i^*). \end{aligned}$$

Part I

On the First Default and Afterwards

Chapter 1

Study on a general framework of credit modelling

The modelling of default time is the key problem for the pricing and the risk management of credit derivatives. Generally speaking, two principal approaches, the structural approach and the reduced form approach, are proposed to model the default mechanism. One important difference between the two approaches is the predictability of the default time with respect to the background filtration.

The structural approach is based on the work of Merton [63] where a firm defaults when its asset value process, often supposed to be represented in terms of a Brownian motion, passes below a certain threshold. This approach provides suitable financial interpretation. However, since the default time is predictable when the asset process approaches the barrier, its disadvantage is also obvious from the modelling point of view. The reduced form approach allows for more "randomness" of the default time. The initial idea comes from the reliability theory where the default is modelled as the first jump of a point process. In this case, the default time is not predictable. The general framework of the reduced form approach has been presented in Jeanblanc and Rutkowski [55] and Elliott, Jeanblanc and Yor [31]. The notion of intensity have often been discussed in this approach. From the mathematical point of view, the intensity process of a stopping time is related to its compensator process, which is a basic notion in the general theory of processes developed in the 1970's. When the compensator satisfies some regularity conditions, the intensity process exists. However, this is not the case in general (see for example Giesecke [34], Guo and Zeng [46]). The gap between the structural approach and the reduced-form approach has been shortened by recent studies of Duffie and Lando [28], Çetin, Jarrow, Protter and Yıldırım [14], Collin-Dufresne Goldstein and Helwege [20], Jeanblanc and Valchev [56] and Guo, Jarrow and Zeng [45], etc. on the impact of information modelling. By specifying certain partial observation hypotheses on the filtrations in the structural approach based models, the default is no longer predictable and the intensity can be calculated in these models.

In this chapter, we review some of the results in the intensity approach from a more general point of view based on the general theory of processes. For theoretical background, one can refer to Dellacherie and Meyer [22], Jacod [49], Protter [67]. However, our objective here is to reinterpret some existing results in the credit modelling using the classical notions without entering in the theoretical details. We are in particular interested in the compensator process of a stopping time τ and in its calculation. An important hypothesis, the (H)-hypothesis, is discussed and a classical example where this hypothesis holds is revisited along this chapter. Instead of the minimal filtration expansion condition often adopted in the credit modelling, we propose to work with a more general condition presented in Jeulin and Yor [57] and discussed in the credit case by Guo, Jarrow and Menn [44], which can be adapted directly to study the first default time in the multi-credits case. Then we are interested in the pricing of the defaultable zero coupon and hence in the calculation of the conditional survival probabilities with respect to the global filtration. To this end, we study the multiplicative decomposition of the so-called "survival process" which is a supermartingale. This leads to a generalization of the classical HJM type model discussed in Schönbucher [72] and we show that the survival process is the key term in determining the conditional survival probabilities. For pricing purposes, especially when there are several credits, it's important to study the case after the default. This is the main issue of the last subsection which consists our main original contribution. We propose a systematic method to calculate the conditional expectations and we point out that the family of conditional survival probabilities $\mathbb{P}(\tau > \theta | \mathcal{F}_t)$ where $\theta \geq 0$ plays the crucial role. We discuss successively the special case where the (H)-hypothesis holds, the case where the survival probability admits a density, and then the general case.

1.1 Stopping time and intensity process: a general framework

We first summarize some general definitions and properties on stochastic processes which are very useful in the following.

1.1.1 Preliminary tools on predictable processes

In the following, let $(\Omega, \mathcal{G}, \mathbb{P})$ be a complete probability space and $\mathbb{G} = (\mathcal{G}_t)_{t\geq 0}$ be a filtration of $\mathcal{G}$ satisfying the usual conditions, that is, the filtration $\mathbb{G}$ is right-continuous and $\mathcal{G}_0$ contains all null sets of $\mathcal{G}_\infty$. The probability space represents the market and the filtration represents the global market information.

The basic results on optional or predictable processes and their dual counterpart may be found in Dellacherie and Meyer [22].

Recall that the $\mathbb{G}$-*predictable σ-algebra* is the σ-algebra $\mathcal{P}^{\mathbb{G}}$ on $\mathbb{R}_+ \times \Omega$ generated by the following processes: $1\!\!1_{\{0\}\times A}$ where $A \in \mathcal{G}_0$ and $1\!\!1_{]s,\infty[\times A}$ where $A \in \mathcal{G}_s$, or by left-continuous adapted processes. Any $\mathcal{P}^{\mathbb{G}}$-measurable process is said to be $\mathbb{G}$-*predictable*. A stopping time σ is said to be a *predictable stopping time* if the process $(1\!\!1_{\{\sigma\le t\}}, t \ge 0)$ is $\mathbb{G}$-predictable.

It is possible to define a kind of universal conditional expectation with respect to the predictable σ-field in the following sense: for any bounded measurable process X, there exists a $\mathbb{G}$-predictable process ${}^pX$, unique up to undistinguishable sense, such that for any $\mathbb{G}$-predictable stopping time σ,

$$\mathbb{E}[X_\sigma 1\!\!1_{\{\sigma<+\infty\}}] = \mathbb{E}[{}^pX_\sigma 1\!\!1_{\{\sigma<+\infty\}}].$$

The process ${}^pX$ is called the $\mathbb{G}$-*predictable projection* of X.
Now, if A is a right-continuous increasing bounded process (not necessary adapted), then there exists a $\mathbb{G}$-predictable increasing process A^p, unique in undistinguishable sense, such that for any bounded measurable process X, we have

$$\mathbb{E}\Big[\int_{[0,+\infty[} X_s dA_s^p\Big] = \mathbb{E}\Big[\int_{[0,+\infty[} {}^pX_s dA_s\Big].$$

The process A^p is called the *dual* $\mathbb{G}$-*predictable projection* of A. In particular, if X is $\mathbb{G}$-predictable, then

$$\mathbb{E}\Big[\int_{[0,+\infty[} X_s dA_s^p\Big] = \mathbb{E}\Big[\int_{[0,+\infty[} X_s dA_s\Big].$$

If A is $\mathbb{G}$-adapted, the process A^p is also the unique increasing predictable process such that the process $A - A^p$ is a $\mathbb{G}$-martingale (cf. [22]).

1.1.2 Compensator process of a stopping time

The first step in the intensity approach of the credit modelling is to precisely define the notion of the *intensity process* of a default time. Since we are concerned with the multi-credits framework, the variable of interest is not necessarily the default time of one credit, but for example the first default time of a portfolio of credits. It is enough to assume that this time is a $\mathbb{G}$-stopping time. In the following, we consider only stopping times which are strictly positive and finite.

Definition 1.1.1 Let τ $(0 < \tau < +\infty)$ be a finite $\mathbb{G}$-stopping time. The $\mathbb{G}$-*compensator* Λ of τ is the dual $\mathbb{G}$-predictable projection of the $\mathbb{G}$-adapted process $(1\!\!1_{\{\tau\le t\}}, t \ge 0)$. The predictable process Λ is also characterized by the martingale property: Λ is a predictable process such that the process $(N_t = 1\!\!1_{\{\tau\le t\}} - \Lambda_t, t \ge 0)$ is a $\mathbb{G}$-martingale. If Λ is absolutely continuous, the $\mathbb{G}$-predictable process λ such that $\Lambda_t = \int_0^t \lambda_s ds$ is called the $\mathbb{G}$-*intensity process* ([31], p.180).

We have the following useful properties of Λ:

1) For any non-negative or bounded $\mathbb{G}$-predictable process H, we have $\mathbb{E}[H_\tau] = \mathbb{E}\big[\int_{[0,\infty[} H_s d\Lambda_s\big]$.
2) The process Λ is stopped at time τ, i.e. $\Lambda_t = \Lambda_{t\wedge\tau}$.
   To see that, we observe that the process $(1\!\!1_{\{\tau\le t\}}, t \ge 0)$ is stopped at τ, and that $(1\!\!1_{\{\tau\le t\}} - \Lambda_{t\wedge\tau}, t \ge 0)$ is a martingale. Then both predictable increasing processes $(\Lambda_t, t \ge 0)$ and $(\Lambda_{t\wedge\tau}, t \ge 0)$ are **undistinguishable**.
3) The process Λ is continuous if and only if τ is a totally inaccessible $\mathbb{G}$-stopping time ([23] p.151), that is, for any $\mathbb{G}$-predictable stopping time σ, $\mathbb{P}(\tau = \sigma) = 0$.
   The jumps of Λ occur at predictable stopping times u such that $\Delta\Lambda_u \le 1$. Moreover, if $\Delta\Lambda_u = 1$, then $u = \tau$, *a.s.*.
4) The $\mathbb{G}$-survival process S, defined by $S_t = 1\!\!1_{\{\tau>t\}} = 1 - 1\!\!1_{\{\tau\le t\}}$, is a right-continuous supermartingale, which satisfies the following equation
   $$dS_t = S_{t-}(-dN_t - d\Lambda_t)$$
   since $S_{t-} = 1\!\!1_{\{\tau\ge t\}}$. This equation will be discussed later.

1.2 Classical frameworks with closed formulae for the compensator processes

One important objective now is to calculate the compensator process Λ of τ. We now present two examples where this computation is done in an explicit way.

1.2.1 The smallest filtration generated by τ

In this example, the filtration is generated by the process associated with random variable $0 < \tau < \infty$. This is a classical but important case which has been largely studied in the single credit modelling (e.g. Elliott, Jeanblanc and Yor [31]) where τ represents the default time of one credit.
Let $\mathbb{D} = (\mathcal{D}_t)_{t\ge 0}$ be the usual augmentation ([22] p.183) of the right-continuous filtration generated by the process $(1\!\!1_{\{\tau\le t\}}; t \ge 0)$. Then $\mathbb{D}$ is the smallest filtration satisfying the usual conditions such that τ is a $\mathbb{D}$-stopping time. Any random variable X is $\mathcal{D}_t$-measurable if and only if

$$X = \widetilde{x} 1\!\!1_{\{\tau>t\}} + x(\tau) 1\!\!1_{\{\tau\le t\}} \tag{1.1}$$

where $\widetilde{x}$ is a constant and x is some Borel function. In particular, the restriction of $\mathcal{D}_t$ on $\{\tau > t\}$ is trivial. Predictable processes, stopped at time τ are deterministic function of $t \wedge \tau$.
To explicitely calculate $\Lambda^{\mathbb{D}}$, the $\mathbb{D}$-compensator of τ, we use the cumulative distribution function F of τ, i.e. $F(t) = \mathbb{P}(\tau \le t)$ and the survival function $G(t) = \mathbb{P}(\tau > t)$.

The following property has been discussed in [31] and [55]:

Proposition 1.2.1 *Let r be the first time such that $G(r) = 0$ (or $F(r) = 1$). For any sufficiently small positive real number ϵ, we have on $[0, \tau \wedge (r - \epsilon)]$,*

$$d\widetilde{\Lambda}_s^{\mathbb{D}} = \frac{dF(s)}{1 - F(s-)} = -\frac{dG(s)}{G(s-)}.$$

Proof. By definition of the compensator process, for any bounded Borel function h,

$$\mathbb{E}\Big[h(\tau)\mathbb{1}_{[0,\tau\wedge(r-\epsilon)]}(\tau)\Big] = \mathbb{E}\Big[\int_{[0,r-\epsilon]} h(s)\mathbb{1}_{\{\tau\geq s\}} d\Lambda_s^{\mathbb{D}}\Big] \tag{1.2}$$

$$= \int_{[0,r-\epsilon]} h(s)dF(s) = \int_{[0,r-\epsilon]} h(s)G(s-)\frac{dF(s)}{G(s-)} = \mathbb{E}\Big[\int_{[0,r-\epsilon]} h(s)\mathbb{1}_{[0,\tau]}(s)\frac{dF(s)}{G(s-)}\Big],$$

The last equality is because $G(s-) = \mathbb{P}(\tau \geq s)$. □

1.2.2 Conditional independance and (H)-hypothesis

An example of stopping time

An important example of default time is given below, which has been discussed by many authors (see Lando [59], Schönbucher and Schubert [74] or Bielecki and Rutkowski [9] for example). The financial interpretation of this model is from the structural approach, at the same time, by introducing a barrier which is independent with the filtration generated by the background process, the intensity of the default time can be calculated. So the two credit modelling approaches are related in this model. Moreover, it has become a standard construction of the default time when given an $\mathbb{F}$-adapted process where $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is an arbitrary filtration on the probability space $(\Omega, \mathcal{G}, \mathbb{P})$.

Let $(\Omega, \mathcal{G}, \mathbb{P})$ be a probability space and $\mathbb{F}$ be a filtration of $\mathcal{G}$. Let Φ be an $\mathbb{F}$-adapted, continuous, increasing process with $\Phi_0 = 0$ and $\Phi_\infty := \lim_{t\to+\infty} \Phi_t = +\infty$. Let ξ be a $\mathcal{G}$-measurable random variable following exponential law with parameter 1 which is independent of $\mathcal{F}_\infty$. We define the random time τ by

$$\tau = \inf\{t \geq 0 : \Phi_t \geq \xi\}.$$

Then we can rewrite τ as $\tau = \inf\big\{t \geq 0 : e^{-\Phi_t} \leq U\big\}$ where $U = e^{-\xi}$ is a uniform random variable on $[0, 1]$. So, the conditional distribution of τ given $\mathcal{F}_\infty$ is given by

$$\mathbb{P}(\tau > t|\mathcal{F}_\infty) = \mathbb{P}(\Phi_t \leq \xi|\mathcal{F}_\infty) = e^{-\Phi_t} = \mathbb{P}(\tau > t|\mathcal{F}_t) =: G_t \tag{1.3}$$

Let us introduce the new filtration $\mathbb{G} = (\mathcal{G}_t)_{t\geq 0}$ as the minimal extension of $\mathbb{F}$ for which τ is a stopping time, that is $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{D}_t$, where $\mathcal{D}_t = \sigma(\mathbb{1}_{\{\tau\leq s\}}, s \leq t)$. In particular, any $\mathcal{G}_t$-measurable random variable coincides with an $\mathcal{F}_t$-measurable r.v. on $\{\tau > t\}$. Based on this observation, we obtain the following characterization of the $\mathbb{G}$-compensator process of τ. Note that Φ is not necessarily continuous here.

Corollary 1.2.2 *Let $\mathbb{G} = \mathbb{F} \vee \mathbb{D}$ be the filtration previously defined. Let us assume that the conditional distribution of τ given $\mathcal{F}_\infty$ is given by $\mathbb{P}(\tau > t|\mathcal{F}_\infty) = e^{-\Phi_t}$ where Φ is an $\mathbb{F}$-adapted increasing process. Then, the $\mathbb{G}$-compensator of τ is the process*

$$\Lambda_s^{\mathbb{G}} = \int_0^{t\wedge\tau} e^{\Phi_{t-}} d(-e^{-\Phi_t})$$

When Φ is differentiable, with derivative $\lambda_t = \partial_t \Phi_t$, then $d\Lambda_t^{\mathbb{G}} = \mathbb{1}_{\{\tau \geq t\}} \lambda_t dt$.

Proof. Since on $\{\tau > t\}$, any $\mathcal{G}_t$-measurable random variable coincides with a $\mathcal{F}_t$-measurable random variable, the $\mathbb{G}$-martingale property of $N_t = \mathbb{1}_{\{\tau \leq t\}} - \Lambda_t^{\mathbb{G}}$ may be expressed as, for any $A_t^{\mathbb{F}} \in \mathcal{F}_t$,

$$\mathbb{E}\Big[\mathbb{1}_{A_t^{\mathbb{F}}}(N_{t+h} - N_t)\Big] = 0 = \mathbb{E}\Big[\mathbb{1}_{A_t^{\mathbb{F}}}(\mathbb{1}_{\{t<\tau\leq t+h\}} - \int_t^{t+h} \mathbb{1}_{\{t<s\leq\tau\}} d\Lambda_s^{\mathbb{G}})\Big].$$

Using the conditional distribution of τ given $\mathcal{F}_\infty$, we have

$$\mathbb{E}\Big[\mathbb{1}_{A_t^{\mathbb{F}}} \mathbb{1}_{\{t<\tau\leq t+h\}}\Big] = \mathbb{E}\Big[\mathbb{1}_{A_t^{\mathbb{F}}} \int_t^{t+h} d(-e^{-\Phi_t})\Big].$$

In order to reintroduce the indicator function $\mathbb{1}_{\{\tau\geq t\}}$ whose $\mathcal{F}_\infty$ conditional expectation is $e^{-\Phi_{t-}}$, we rewrite the right-hand side of the above equality as

$$\mathbb{E}\Big[\mathbb{1}_{A_t^{\mathbb{F}}} \int_t^{t+h} d(-e^{-\Phi_t})\Big] = \mathbb{E}\Big[\mathbb{1}_{A_t^{\mathbb{F}}} \int_t^{t+h} \mathbb{1}_{\{\tau\geq s\}} e^{\Phi_{t-}} d(-e^{-\Phi_t})\Big]$$

By uniqueness of the dual predictable decomposition, we know that the corollary holds. □

The process G defined by $G_t = \mathbb{P}(\tau > t|\mathcal{F}_t)$ is called the $\mathbb{F}$-*survival process* . In this framework, G is a decreasing $\mathbb{F}$-adapted process. However, this is not true in general, as we will see in the following section. We note that this example is from the "filtration expansion" point of view, that is, the filtration $\mathbb{G}$ is set to be the "minimal expansion" of $\mathbb{F}$ as been discussed in Mansuy and Yor [62].

About the (H)-Hypothesis.

We now introduce the (H)-hypothesis first introduced in filtering theory by many authors, see Brémaud and Yor [12] for instance.

Hypothesis 1.2.3 We say that the (H)-*hypothesis* holds for $(\mathbb{F}, \mathbb{G})$, or equivalently, we say that a sub-filtration $\mathbb{F}$ of $\mathbb{G}$ has the *martingale invariant property* with respect to the filtration $\mathbb{G}$ or that $\mathbb{F}$ is *immersed* in $\mathbb{G}$, if any $\mathbb{F}$ square-integrable martingale is a $\mathbb{G}$-martingale.

For the credit modelling purpose, it has been studied by Kusuoka [58] and Jeanblanc and Rutkowski [55]. Note that the (H)-hypothesis is often supposed in the credit modelling, since it is equivalent to conditional independence between the σ-algebras $\mathcal{G}_t$ and $\mathcal{F}_\infty$ given $\mathcal{F}_t$ (see [22]) as in the above example.

We give some equivalent forms below (cf. [22]). The (H)-hypothesis for $(\mathbb{F}, \mathbb{G})$ is equivalent to any of the following conditions:

(H1) for any $t \geq 0$ and any bounded $\mathcal{G}_t$-measurable r.v. Y we have $\mathbb{E}[Y|\mathcal{F}_\infty] = \mathbb{E}[Y|\mathcal{F}_t]$;

(H2) for any $t \geq 0$, any bounded $\mathcal{G}_t$-measurable r.v. Y and any bounded $\mathcal{F}_\infty$-measurable r.v. Z, we have $\mathbb{E}[YZ|\mathcal{F}_t] = \mathbb{E}[Y|\mathcal{F}_t]\mathbb{E}[Z|\mathcal{F}_t]$.

Remark 1.2.4 The main disadvantage of the (H)-hypothesis is that it may fail to hold under a change of probability. Kusuoka [58] provided a counter example to show this property. One can also refer to Bielecki and Rutkowski [9] and Jeanblanc and Rutkowski [55] for a detailed review.

1.3 General framework for credit modelling with two filtrations

In this subsection, we present our framework which is an extension of the classical frameworks in the credit modelling. Recall that the filtration $\mathbb{G}$ represents the global information on the market. We now make our main assumption introduced by Jeulin and Yor [57] in 1978 and Jacod [49] in 1979. This assumption has been discussed recently in Guo, Jarrow and Menn [44] and Guo and Zeng [46] (2006) as a general filtration expansion condition. We are from another point of view. Our reference filtration is the global one $\mathbb{G}$.

Main Assumption

Hypothesis 1.3.1 (Minimal Assumption) Let $\mathbb{F}$ be a subfiltration of the general filtration $\mathbb{G}$ and let τ $(0 < \tau < \infty)$ be a $\mathbb{G}$-stopping time. We say that $(\mathbb{F}, \mathbb{G}, \tau)$ satisfy the Minimal Assumption (MA) if for any $t \geq 0$ and any $U \in \mathcal{G}_t$, there exists $V \in \mathcal{F}_t$ such that

$$U \cap \{\tau > t\} = V \cap \{\tau > t\}. \tag{1.4}$$

Obviously the filtrations introduced in the previous example, $\mathbb{F}$ and $\mathbb{G} = \mathbb{F} \vee \mathbb{D}$, verify this assumption, since the filtration $\mathbb{D}$ is trivial on $\{\tau > t\}$. The Hypothesis 1.3.1 and the framework we shall introduce is quite general, and well-adapted to study what happen before the first default time in the multi-credits case. We shall show that the first default time can be treated in almost the same way as for one default time in the

single-credit case.
We first give some consequences of the Minimal Assumption on $\mathbb{G}$-predictable processes in [57].

Proposition 1.3.2 *Assume that $\mathcal{G}_0 = \mathcal{F}_0$, and that $(\mathbb{F}, \mathbb{G}, \tau)$ satisfy the Minimal Assumption 1.3.1.*
Then for any $\mathbb{G}$-predictable process $H^{\mathbb{G}}$, there exists an $\mathbb{F}$-predictable process $H^{\mathbb{F}}$ such that

$$H_t^{\mathbb{G}} 1\!\!1_{\{\tau \geq t\}} = H_t^{\mathbb{F}} 1\!\!1_{\{\tau \geq t\}}. \tag{1.5}$$

Proof. The $\mathbb{G}$-predictable σ-algebra on $[0, \infty[\times\Omega$ is generated by the following two types of subsets: $\{0\} \times A$ with $A \in \mathcal{G}_0$ and $]s, \infty] \times A$ with $A \in \mathcal{G}_r$, $r < s$. For processes such that $H^{\mathbb{G}} = 1\!\!1_{\{0\}\times A}$, the property holds automically since the hypothesis $\mathcal{F}_0 = \mathcal{G}_0$ implies that $H^{\mathbb{G}}$ is $\mathbb{F}$-predictable.
We now need to prove (1.5) when $H^{\mathbb{G}} = 1\!\!1_{]s,\infty]\times A}$ for any $s \geq 0$ and any $A \in \mathcal{G}_s$. We know that there exists $A^{\mathbb{F}} \in \mathcal{F}_s$ such that $A \cap \{\tau > s\} = A^{\mathbb{F}} \cap \{\tau > s\}$. Let $H^{\mathbb{F}} = 1\!\!1_{]s,\infty]\times A^{\mathbb{F}}}$. Then, we have (1.5).
□

$\mathbb{F}$-survival process and $\mathbb{G}$-compensator

The above proposition enables us to calculate the process $H^{\mathbb{F}}$. In fact, by taking conditional expectations, we have

$$H_t^{\mathbb{F}} \mathbb{P}(\tau \geq t | \mathcal{F}_{t-}) = \mathbb{E}[H_t^{\mathbb{G}} 1\!\!1_{\{\tau \geq t\}} | \mathcal{F}_{t-}]. \tag{1.6}$$

As in the example, the $\mathbb{F}$-conditional probability $\mathbb{P}(\tau \geq t|\mathcal{F}_{t-})$ plays an important role in different calculations.

Definition 1.3.3 The $\mathbb{F}$-*survival process* G is the right continuous supermartingale

$$G_t = \mathbb{P}(\tau > t|\mathcal{F}_t). \tag{1.7}$$

The point 0 is absorbing for this nonnegative supermartingale.

Note that G is the *Azéma supermartingale* ([62], p.13) of τ.
The following theorem enables us to calculate the $\mathbb{G}$-compensator process via the process G. The first result is a classical one which has been given in [31]. The second result is an extension which shall be useful in the multi-credits case.

Theorem 1.3.4 *We assume that Hypothesis 1.3.1 holds.*

1) Let $G = M - A$ be the Doob-Meyer decomposition of the survival process G, where M is a martingale and A is an increasing $\mathbb{F}$-predictable process.
On $\{G_{t-} > 0\}$, we have the following representation of the $\mathbb{G}$-compensator process $\Lambda^{\mathbb{G}}$ of τ.

$$d\Lambda_t^{\mathbb{G}} = 1\!\!1_{]0,\tau]}(t)\frac{dA_t}{G_{t-}} = 1\!\!1_{]0,\tau]}d\Lambda_t^{\mathbb{F}} \tag{1.8}$$

2) More generally, let σ be another $\mathbb{G}$-stopping time with $\mathbb{G}$-compensator Λ^σ. Then there exists an increasing $\mathbb{F}$-predictable process $\Lambda^{\sigma,\mathbb{F}}$ such that $\Lambda^\sigma_{t\wedge\tau} = \Lambda^{\sigma,\mathbb{F}}_{t\wedge\tau}$. Let $B^{\sigma,\mathbb{F}}$ be the $\mathbb{F}$-predictable increasing process in the Doob-Meyer decomposition of the supermartingale $V_t^\sigma = \mathbb{P}(\sigma > \tau \wedge t|\mathcal{F}_t)$, or in other words, $V^\sigma + B^{\sigma,\mathbb{F}}$ is an $\mathbb{F}$-martingale. Then

$$dB_t^{\sigma,\mathbb{F}} = G_{t-}d\Lambda_t^{\sigma,\mathbb{F}}$$

Proof. 1) The $\mathbb{G}$-predictable process Λ is stopped at τ, i.e. $\Lambda_t = \Lambda_{t\wedge\tau}$. Hence, under hypothesis 1.3.1, there exists an $\mathbb{F}$-predictable process $\Lambda^{\mathbb{F}}$ such that $\Lambda_t = \Lambda_{t\wedge\tau} = \Lambda^{\mathbb{F}}_{t\wedge\tau}$. On the other hand, Λ admits a dual $\mathbb{F}$-predictable projection, which coincides with the $\mathbb{F}$-predictable increasing process A in the Doob-Meyer decomposition of G. Since we have supposed that $\tau < +\infty$ a.s., for any bounded $\mathbb{F}$-predictable process Y,

$$\mathbb{E}[Y_\tau] = \mathbb{E}\Big[\int_{[0,+\infty[} Y_t d\Lambda_t\Big] = \mathbb{E}\Big[\int_{[0,+\infty[} Y_t dA_t\Big].$$

On the other side, since $\Lambda_t = \Lambda^{\mathbb{F}}_{\tau\wedge t}$ by hypothesis, we know that

$$\mathbb{E}\Big[\int_{[0,+\infty[} Y_t d\Lambda_t\Big] = \mathbb{E}\Big[\int_{[0,+\infty[} 1\!\!1_{\{\tau\geq t\}} Y_t d\Lambda_t^{\mathbb{F}}\Big] = \mathbb{E}\Big[\int_{[0,+\infty[} Y_t G_{t-} d\Lambda_t^{\mathbb{F}}\Big].$$

The last equality is because $G_{t-} = \mathbb{E}[1\!\!1_{\{\tau\geq t\}}|\mathcal{F}_t]$. Since Y is arbitrary, we know that $dA_t = G_{t-}d\Lambda_t^{\mathbb{F}}$.

2) By definition, $(\Lambda^\sigma_{\tau\wedge t}, t \geq 0)$ is the compensator of the process $(1\!\!1_{\{\sigma\leq\tau\wedge t\}}, t \geq 0)$. Therefore, for any bounded $\mathbb{F}$-predictable process Y,

$$\mathbb{E}\Big[\int_{[0,+\infty[} Y_t d1\!\!1_{\{\sigma\leq\tau\wedge t\}}\Big] = \mathbb{E}\Big[\int_{[0,+\infty[} Y_t d\Lambda^\sigma_{\tau\wedge t}\Big] = -\mathbb{E}\Big[\int_{[0,+\infty[} Y_t dV_t^\sigma\Big] = \mathbb{E}\Big[\int_{[0,+\infty[} Y_t dB_t^{\sigma,\mathbb{F}}\Big].$$

The last two equalities come from the definition of V^σ and $B^{\sigma,\mathbb{F}}$. Since $\Lambda^\sigma_{\tau\wedge t} = \Lambda^{\sigma,\mathbb{F}}_{\tau\wedge t}$, similarly as in 1), we have $\mathbb{E}\big[\int_{[0,+\infty[} Y_t d\Lambda^{\sigma,\mathbb{F}}_{\tau\wedge t}\big] = \mathbb{E}\big[\int_{[0,+\infty[} Y_t G_{t-} d\Lambda_t^{\sigma,\mathbb{F}}\big]$, which ends the proof.
□

Multiplicative decomposition of G

In addition to the additive decomposition of the $\mathbb{F}$-survival process G which is an $\mathbb{F}$-supermartingale, there exists a multiplicative decomposition of G, very useful in the following. This property has been discussed in Jeanblanc and LeCam [53]. For simplicity, we introduce the following assumption.

Hypothesis 1.3.5 We assume that $G_t > 0$ a.s. for any $t \geq 0$.

However, certain results hold without this hypothesis before the first time that G attains zero as shown by Theorem 1.3.4.

To deduce the explicit form of the multiplicative decomposition of G, we introduce the *Doléans-Dade exponential of a semimartingale* Z which is the unique solution of the stochastic differential equation

$$dX_t = X_{t-} dZ_t, \quad X_0 = 1,$$

which is given by

$$\mathcal{E}(Z)_t = \exp\left(Z_t - Z_0 - \frac{1}{2} \langle Z \rangle_t^c\right) \prod_{0<s\leq t} \Big[\big(1 + \Delta Z_s\big) e^{-\Delta Z_s}\Big]. \tag{1.9}$$

Observe that if Z is a local martingale, then $\mathcal{E}(Z)$ is also a local martingale, and that if B is a predictable increasing process such that $\Delta B < 1$, then $\mathcal{E}(-B)$ is a predictable decreasing process. Note that the Doléans-Dade exponential does not satisfy the standard exponential calculation property. In fact, for any semimartingales X and Y, we have $\mathcal{E}(X + Y + [X, Y]) = \mathcal{E}(X)\mathcal{E}(Y)$.
We are looking for a multiplicative decomposition of the positive supermartingale G as

$$G = L\, D \tag{1.10}$$

where L is an $\mathbb{F}$-local martingale and D is an $\mathbb{F}$-predictable decreasing process.
We start with the Doob-Meyer decomposition of G as $G = M - A$ where M is an $\mathbb{F}$-local martingale and A is an $\mathbb{F}$-predictable increasing process. Then we have

$$\frac{dG_s}{G_{s-}} = \frac{dM_s}{G_{s-}} - \frac{dA_s}{G_{s-}} = dM_s^{\widetilde{\Gamma}} - d\Lambda_s^{\mathbb{F}}$$

where $dM^{\widetilde{\Gamma}} := dM/G_-$. Assume $\Lambda^{\mathbb{F}}$ to be **continuous**. Then, the additivity property holds for the Doléans-Dade exponentials of the processes $M^{\widetilde{\Gamma}}$ and $-\Lambda_s^{\mathbb{F}}$, that is,

$$G = \mathcal{E}(M^{\widetilde{\Gamma}} - \Lambda^{\mathbb{F}}) = \mathcal{E}(M^{\widetilde{\Gamma}}) \exp(-\Lambda^{\mathbb{F}})$$

and the multiplicative decomposition is obvious.
In the general case, the problem is much more complicated and less intuitive. The good answer, as proposed by Jacod (Corollary 6.35 in [49]), Meyer [64] and Yoeurp [81] is given in the following proposition.

Proposition 1.3.6 *Let* $\widetilde{\Gamma}_t = \ln G_0 + \int_0^t \frac{dG_s}{G_{s-}}$ *and* $\widetilde{\Gamma}_t = M_t^{\widetilde{\Gamma}} - \Lambda_t^{\mathbb{F}}$ *be its Doob-Meyer decomposition. Assume that* $\Delta\Lambda^{\mathbb{F}} \neq 1$.
Then G *admits a multiplicative decomposition as*

$$G = \mathcal{E}(\widetilde{\Gamma}) = \mathcal{E}(\widetilde{M^{\widetilde{\Gamma}}})\mathcal{E}(-\Lambda^{\mathbb{F}}) \tag{1.11}$$

where the $\mathbb{F}$*-local martingale* $\widetilde{M^{\widetilde{\Gamma}}}$ *is defined by* $d\widetilde{M_t^{\widetilde{\Gamma}}} = \frac{1}{1-\Delta\Lambda_t^{\mathbb{F}}} dM_t^{\widetilde{\Gamma}}$.
In other words, the martingale part L *of the multiplicative decomposition of* G *is the Doléans-Dade exponential of the modified martingale* $\widetilde{M^{\widetilde{\Gamma}}}$ *and the predictable decreasing process* D *is the Doléans-Dade exponential of* $-\Lambda^{\mathbb{F}}$.

Proof. It suffices to check that the right hand side of (1.11) is the needed decomposition. We know that $\mathcal{E}(\widetilde{M^{\widetilde{\Gamma}}})\mathcal{E}(-\Lambda^{\mathbb{F}}) = \mathcal{E}\big(\widetilde{M^{\widetilde{\Gamma}}} - \Lambda^{\mathbb{F}} + [\widetilde{M^{\widetilde{\Gamma}}}, -\Lambda^{\mathbb{F}}]\big)$ and we shall prove that $[\widetilde{M^{\widetilde{\Gamma}}}, -\Lambda^{\mathbb{F}}]$ is a pure jump martingale. It suffices to prove that for any predictable process $Z^{\mathbb{F}}$, $I = \mathbb{E}\big[\sum_s Z_s^{\mathbb{F}} \Delta\widetilde{M_s^{\widetilde{\Gamma}}} \Delta\Lambda_s^{\mathbb{F}}\big] = 0$. Recall that the $\mathbb{F}$-predictable projection of the martingale $\widetilde{M^{\widetilde{\Gamma}}}$ is $\widetilde{M_-^{\widetilde{\Gamma}}}$ and that $\widetilde{M_-^{\widetilde{\Gamma}}}$ has no predictable jumps. Since $\Lambda^{\mathbb{F}}$ is a increasing predictable process, we know that $I = \mathbb{E}[\sum_s Z_s^{\mathbb{F}} (\widetilde{M_{s-}^{\widetilde{\Gamma}}} - \widetilde{M_{s-}^{\widetilde{\Gamma}}}) \Delta\Lambda_s^{\mathbb{F}}] = 0$. Moreover, the other purely jumps martingale $N = M^{\widetilde{\Gamma}} - \widetilde{M^{\widetilde{\Gamma}}} + [\widetilde{M^{\widetilde{\Gamma}}}, \Lambda^{\mathbb{F}}]$ has no jump since

$$\Delta N = \Delta M^{\widetilde{\Gamma}} \big(1 - \frac{1}{1-\Delta\Lambda^{\mathbb{F}}} + \frac{\Delta\Lambda^{\mathbb{F}}}{1-\Delta\Lambda^{\mathbb{F}}}\big) = 0$$

Therefore, the martingale N is identically equal to 0. That ends the proof.

□

1.4 Defaultable zero coupon and conditional survival probabilities

We are now interested in the evaluation of the *defaultable zero coupon*. A defaultable zero coupon is the financial product where an investor receives 1 monetary unit at maturity T if no default occurs before T and 0 otherwise. Here, we don't take into account the discounting impact and the risk-neutral evaluation. Hence the conditional probability $\mathbb{P}(\tau > T|\mathcal{G}_t)$ is the key term to evaluate. By previous discussions, we know that this $\mathbb{G}$-martingale coincides on the set $\{\tau > t\}$ with some $\mathbb{F}$-adapted process and it can be calculated by

$$\mathbb{P}(\tau > T|\mathcal{G}_t) = \mathbb{1}_{\{\tau>t\}} \frac{\mathbb{P}(\tau > T|\mathcal{F}_t)}{\mathbb{P}(\tau > t|\mathcal{F}_t)} = \mathbb{1}_{\{\tau>t\}} \frac{\mathbb{E}[G_T|\mathcal{F}_t]}{G_t} = \mathbb{1}_{\{\tau>t\}} \widetilde{G}_t^T,$$

where $\widetilde{G}_t^T = \mathbb{E}[G_T|\mathcal{F}_t]/G_t$. Here the intensity is not the suitable tool to study the problem.

1.4.1 General Framework and abstract HJM

The multiplicative decomposition of $G = L\,D$ in terms of the $\mathbb{F}$-local martingale $L = \mathcal{E}(\widetilde{M^\Gamma})$ and the $\mathbb{F}$-predictable decreasing process $D = \mathcal{E}(-\Lambda^{\mathbb{F}})$ is the natural tool to study $\widetilde{G}_t^T$ under the following assumption:

Hypothesis 1.4.1 Assume that $\Delta\Lambda^{\mathbb{F}} \neq 1$, and that the exponential martingale $L = \mathcal{E}(\widetilde{M^\Gamma})$ is strictly positive on $[0,T]$. Then, the change of probability measure

$$d\mathbb{Q}^L = L_T\,d\mathbb{P} \quad \text{on} \quad \mathcal{F}_T \tag{1.12}$$

is well-defined.

Remark 1.4.2 When the (H)-hypothesis holds between the filtrations $\mathbb{F}$ and $\mathbb{G}$, the $\mathbb{F}$-survival process G is the exponential of an adapted decreasing process $-\Phi$. So we can work directly under the initial probability measure $\mathbb{P}$, that is $L_T = 1$.

Forward hazard process

We now study properties of $\widetilde{G}_t^T$ in both directions, as a function of T or as a semi-martingale on t. Given the multiplicative decomposition of G, we have

$$\widetilde{G}_t^T = \frac{\mathbb{E}\big[G_T|\mathcal{F}_t\big]}{G_t} = \frac{\mathbb{E}\big[L_T\,D_T|\mathcal{F}_t\big]}{L_t\,D_t} = \mathbb{E}_{\mathbb{Q}^L}\big[D_{T,t}|\mathcal{F}_t\big]$$

where $D_{T,t} = \frac{D_T}{D_t}$ is $\mathcal{F}_T$-measurable. Since D is the Doléans-Dade exponential of $-\Lambda^{\mathbb{F}}$, under our hypothesis, D is decreasing, and we can introduce another predictable increasing process $\mathcal{L}^{\mathbb{F}}$ such that

$$\exp(-\mathcal{L}_t^{\mathbb{F}}) = \mathcal{E}(-\Lambda_t^{\mathbb{F}}) = D_t.$$

By analogy with zero-coupon modelling, we introduced the process Γ_t^T defined as the parallel of the logarithm of "defaultable zero-coupon bond". This process is known as the *forward hazard process*

$$\Gamma_t^T = -\ln \widetilde{G}_t^T = -\ln \mathbb{E}_{\mathbb{Q}^L}\big[\exp\big(-(\mathcal{L}_T^{\mathbb{F}} - \mathcal{L}_t^{\mathbb{F}})\big)|\mathcal{F}_t\big]. \tag{1.13}$$

Then we have the abstract version of HJM framework for the $\mathbb{G}$-survival probability before the default time.

Proposition 1.4.3 (Abstract HJM) *We take previous notations.*

1) *For any $t \leq T$, the process $\big(\mathbb{P}(\tau > T|\mathcal{G}_t) = \mathbb{1}_{\{\tau>t\}} e^{-\Gamma_t^T}, t \geq 0\big)$ is a $\mathbb{G}$-martingale.*

2) *If Λ is continuous, then the process $\big(\mathbb{1}_{\{\tau>t\}} e^{\Lambda_t} = \mathbb{1}_{\{\tau>t\}} e^{\Lambda^{\mathbb{F}}_{t\wedge\tau}}, t \geq 0\big)$ is a $\mathbb{G}$-martingale.*

3) The process $\widehat{G}^T$ defined by $\widehat{G}_t^T = \exp -(\mathcal{L}_t^{\mathbb{F}} + \Gamma_t^T) = \mathbb{E}_{\mathbb{Q}^L}[D_T|\mathcal{F}_t]$ is a $\mathbb{Q}^L$-martingale on $[0,T]$ with respect to the filtration $\mathbb{F}$.

4) Assume the process $\Lambda^{\mathbb{F}}$ to be absolutely continuous, with $\mathbb{F}$-intensity process λ regular enough to make valid differentiation under $\mathbb{Q}^L$ expectation.
Then the semimartingale process Γ^T is also absolutely continuous with respect to T. Its "derivative" process is the $\mathbb{F}$-forward intensity process $\gamma(t,T)$ such that $\Big(\exp\Big(-\int_0^T \big(\lambda_u 1\!\!1_{[0,t]}(u) + \gamma(t,u)1\!\!1_{[t,T]}(u)\big)du\Big), t \geq 0\Big)$ is an $\mathbb{F}$-martingale on $[0,T]$ with respect to the $\mathbb{Q}^L$ probability measure.

Proof. 1) is obvious by definition.
2) By the multiplicative decomposition, $S_t = 1\!\!1_{\{\tau>t\}} = \mathcal{E}(-\Lambda_t)\mathcal{E}(-N_t)$ where Λ is the $\mathbb{G}$-compensator process of τ and N is the $\mathbb{G}$-martingale in the Doob-Meyer decomposition of S. If Λ is continuous, then $\mathcal{E}(-\Lambda_t) = \exp(-\Lambda_t)$. So $(1\!\!1_{\{\tau>t\}}e^{\Lambda_t} = \mathcal{E}(-N_t), t \geq 0)$ is a $\mathbb{G}$-martingale.
3) is direct by (1.13).
4) Since $\Lambda^{\mathbb{F}}$ is absolutely continuous and is of density λ, $\mathcal{E}(-\Lambda_t^{\mathbb{F}}) = \exp\big(-\int_0^t \lambda_u du\big)$. Then by definition, $\Gamma_t^T = -\ln \mathbb{E}_{\mathbb{Q}^L}\big[e^{-\int_t^T \lambda_u du}\big|\mathcal{F}_t\big]$ is absolutely continuous with respect to T. If we denote by $\gamma(t,T)$ the $\mathbb{F}$-forward intensity process — the derivative of Γ_t^T with respect to T, then the process defined in 3) is nothing but $\widehat{G}^T$, which implies the desired result. □

1.4.2 HJM model in the Brownian framework

As been shown above, the zero coupon prices gives us the necessary information on the $\mathbb{G}$-conditional probability $\mathbb{P}(\tau > T|\mathcal{G}_t)$ for all $T > t$. We now explicate this point in the classical context of the HJM model. In the following of this subsection, we assume that the filtration $\mathbb{F}$ is generated by a $\mathbb{Q}^L$ Brownian motion $\widehat{W}$ and that the $\mathbb{F}$-predictable process $\Lambda^{\mathbb{F}}$ is absolutely continuous. In this case, the process G and its martingale part L are continuous. In addition, we suppose that the Hypothesis 1.3.1 holds for the filtrations $(\mathbb{F},\mathbb{G})$.

HJM approach was first developed by Heath, Jarrow and Morton [47] to describe the dynamics of the term structure of interest rate. Modelling the whole family of interest rate curves for all maturities appears to be a difficult problem with infinite dimension. While in fact under the condition of absence of arbitrage opportunity, the dynamics of the forward rate is totally determined by the short-term rate today and the volatility coefficient. Application of this approach on the credit study is introduced by Jarrow and Turnbull [50], Duffie [26] and Duffie and Singleton [29]. In Schönbucher [72] and Bielecki and Rutkowski [9], one can find related descriptions of this approach applied to the defaultable term structure and defaultable bond of a single credit.

Schönbucher [72] used the HJM framework to represent the term structure of the defaultable bond and give the arbitrage-free conditions. In the following, we proceed from a different point of view. First, we work under the modified probability $\mathbb{Q}^L$. Thus we proceed similarly as in the interest rate modelling. On the other hand, as mentioned above, the only difference when the (H)-hypothesis holds is that there is no need to change the probability. Therefore, we can deal with the general case without extra effort than in the special case with (H)-hypothesis. Second, instead of supposing the dynamics of the forward rate, we here suppose to know the dynamics of the $\mathbb{F}$-martingale $\widehat{G}^T$ under the probability $\mathbb{Q}^L$. This is exactly as in the interest rate modelling where the discounted zero coupon price has a martingale representation. We deduce the HJM model under the probability $\mathbb{Q}^L$. Then it's easy to obtain G_t^T by multiplying the discounted factor D_t. We then deduce the dynamics of the $\mathbb{F}$-survival process G and thus the $\mathbb{G}$-conditional probability $\mathbb{P}(\tau > T|\mathcal{G}_t)$.

Proposition 1.4.4 *Assume that for any $T > 0$, the process $(\widehat{G}_t^T, 0 \leq t \leq T)$ satisfies the following equation:*

$$\frac{d\widehat{G}_t^T}{\widehat{G}_t^T} = \Psi(t,T)d\widehat{W}_t \tag{1.14}$$

where $(\Psi(t,T), t \in [0,T])$ is an $\mathbb{F}$-adapted process which is differentiable with respect to T and $\widehat{W}$ is a Brownian motion under the probability $\mathbb{Q}^L$. If, in addition, $\psi(t,T) = \frac{\partial}{\partial T}\Psi(t,T)$ is bounded uniformly on (t,T), then we have

1)

$$\widehat{G}_t^T = \widehat{G}_0^T \exp\left[\int_0^t \Psi(s,T)d\widehat{W}_s - \frac{1}{2}\int_0^t |\Psi(s,T)|^2 ds\right] \tag{1.15}$$

2)

$$\widehat{\gamma}(t,T) = \widehat{\gamma}(0,T) - \int_0^t \psi(s,T)d\widehat{W}_s + \int_0^t \psi(s,T)\Psi(s,T)^* ds. \tag{1.16}$$

3) We have $\Psi(u,u) = 0$ and

$$\widehat{G}_t = \exp\left[-\int_0^t \widehat{\gamma}(s,s)ds\right]. \tag{1.17}$$

4)

$$\widehat{\gamma}(t,T) = \widehat{\gamma}(T,T) + \int_t^T \psi(s,T)d\widehat{W}_s - \int_t^T \psi(s,T)\Psi(s,T)^* ds. \tag{1.18}$$

Proof. 1) The first equation is the explicit form of the solution of equation (1.14).

2) By definition, $\widehat{\gamma}(t,T)$ is obtained by taking the derivative of $-\ln \widehat{G}_t^T$ with respect to T. Then combining (1.15), we get

$$\begin{aligned}\widehat{\gamma}(t,T) &= -\frac{\partial}{\partial T}\ln \widehat{G}_t^T = -\frac{\partial}{\partial T}\left[\ln \widehat{G}_0^T + \int_0^t \Psi(s,T)d\widehat{W}_s - \frac{1}{2}\int_0^t |\Psi(s,T)|^2 ds\right] \\ &= \widehat{\gamma}(0,T) - \int_0^t \psi(s,T)d\widehat{W}_s + \int_0^t \psi(s,T)\Psi(s,T)^* ds.\end{aligned}$$

3) Equality (1.15) implies that

$$\begin{aligned}\ln \widehat{G}_t &= \ln \widehat{G}_0^t + \int_0^t \Psi(s,t)d\widehat{W}_s - \frac{1}{2}\int_0^t |\Psi(s,t)|^2 ds \\ &= -\int_0^t \widehat{\gamma}(0,s)ds + \int_0^t \Psi(s,t)d\widehat{W}_s - \frac{1}{2}\int_0^t |\Psi(s,t)|^2 ds.\end{aligned} \tag{1.19}$$

Moreover, we have from equation (1.16) that

$$\begin{aligned}&\int_0^t \widehat{\gamma}(s,s)ds \\ &= \int_0^t \widehat{\gamma}(0,s)ds - \int_0^t ds\int_0^s \psi(u,s)d\widehat{W}_u + \int_0^t ds \int_0^s \psi(u,s)\Psi(u,s)^* du \\ &= \int_0^t \widehat{\gamma}(0,s)ds - \int_0^t (\Psi(u,t)-\Psi(u,u))d\widehat{W}_u + \frac{1}{2}(\int_0^t |\Psi(u,t)|^2 - \int_0^t |\Psi(u,u)|^2)du.\end{aligned} \tag{1.20}$$

Combining (1.19) and (1.20) we get

$$\widehat{G}_t = \exp\left[-\int_0^t \widehat{\gamma}(s,s)ds + \int_0^t \Psi(u,u)d\widehat{W}_u - \frac{1}{2}\int_0^t |\Psi(u,u)|^2 du\right].$$

On the other side, $\widehat{G} = D$ is a decreasing process, so its martingale part vanishes, which implies that $\Psi(t,t) = 0$ for any $t \geq 0$.

4) is a direct result from 2). □

Remark 1.4.5 The above result can be applied directly to the first default time in the multi-credits case since the condition we need for the filtrations $(\mathbb{F},\mathbb{G})$ is the general Hypothesis 1.3.1 which is fulfilled in this case.

1.5 After the default event

This section is devoted to the study on the period after the default. The point of view and the results presented below are extension of the work of Bielecki, Jeanblanc and Rutkowski [79]. We have studied, in the previous section, the $\mathbb{G}$-survival probability

$\mathbb{P}(\tau > T|\mathcal{G}_t)$ for all $T \geq t$. It is shown that the knowledge on the process G enables us to calculate this conditional survival probability since it equals zero on the set $\{\tau \leq t\}$. For pricing purposes, we are now interested in the calculation of the general $\mathbb{G}$-conditional expectations. However, we shall distinguish two cases before and after the default in the general case. Before the default, that is, on the set $\{\tau > t\}$, we know from previous discussions that the calculation is easy. Recall that for any $\mathcal{G}_T$-measurable random variable Y, we have $\mathbb{E}[Y|\mathcal{G}_t]1\!\!1_{\{\tau>t\}} = \frac{\mathbb{E}[Y1\!\!1_{\{\tau>t\}}|\mathcal{F}_t]}{\mathbb{P}(\tau>t|\mathcal{F}_t)}1\!\!1_{\{\tau>t\}}$. We observe again the important role played by the process G. In addition, for computation purposes, we need a martingale characterization of the $\mathbb{F}$-martingale of the form $\mathbb{E}[Y1\!\!1_{\{\tau>t\}}|\mathcal{F}_t]$. This is one issue we shall study in this section.

Moreover, if we consider a single default, it suffices to consider the case on $\{\tau > t\}$. That's what many models on the market study. However, to extend our framework to several default times, we have to understand what occurs after the default, that is, on the set $\{\tau \leq t\}$. This is of great importance while studying CDS prices, or k^{th}-to-default products. We begin our discussion by a special case where the (H)-hypothesis holds.

1.5.1 A special case with (H)-hypothesis

We now revisit the example where (H)-hypothesis holds for $(\mathbb{F}, \mathbb{G})$ and we suppose that Φ is absolutely continuous. With the notation of the previous section, for any $T > t$, we have

$$G_t^T := \mathbb{P}(\tau > T|\mathcal{F}_t) = \exp\big(-\int_0^T \big(\lambda_s 1\!\!1_{\{s\leq t\}} + \gamma(t,s)1\!\!1_{\{s>t\}}\big)ds\big).$$

We now consider the conditional probability $\mathbb{P}(\tau > s|\mathcal{F}_t)$ where $s < t$, which is important for the case after the default. By property of the (H)-hypothesis, we have $\mathbb{P}(\tau > t|\mathcal{F}_t) = \mathbb{P}(\tau > t|\mathcal{F}_\infty)$. Hence, for any $T \geq t$, we have $\mathbb{P}(\tau > t|\mathcal{F}_T) = G_t$, which implies immediately that

$$\mathbb{P}(\tau > s|\mathcal{F}_t) = \mathbb{P}(\tau > s|\mathcal{F}_\infty) = G_s = \exp\Big(-\int_0^s \lambda_u du\Big).$$

Combining the two cases, we obtain $\mathbb{P}(\tau > \theta|\mathcal{F}_t) = G_{t\wedge\theta}^\theta$ for any $\theta \geq 0$. This conditional probability admits a density $\alpha_t(\theta)$ given by

$$\alpha_t(\theta) = \begin{cases} \lambda_\theta \exp\big(-\int_0^\theta \lambda_u du\big), & \theta \leq t \\ \gamma(t,\theta)\exp\big(-\int_0^t \lambda_u du - \int_t^\theta \gamma(t,u)du\big), & \theta > t \end{cases} \tag{1.21}$$

such that $\mathbb{P}(\tau > \theta|\mathcal{F}_t) = \int_\theta^\infty \alpha_t(u)du$.

Remark 1.5.1 The density $\alpha_t(\theta)$ does not depend on t for $t \geq \theta$ when the (H)-hypothesis holds, which simplifies sometimes the calculation.

In the evaluation problem, it's very important to calculate the conditional expectation of a $\mathcal{G}_T$-measurable random variable with respect to $\mathcal{G}_t$ where T is the maturity and t is the evaluation date. In the following, we consider the random variable of the form $Y(T,\tau)$ where for any $s \geq 0$, $Y(T,s)$ is $\mathcal{F}_T$-measurable and for any $\omega \in \Omega$, $Y(T,s)$ is a Borel function of s. Notice that any $\mathcal{G}_T$-measurable random variable can be written in this form. However $Y(T,\tau)$ can represent a larger set of random variables. The following result holds for all $Y(T,\tau)$ defined above. However, for pricing purposes, we are only interested in $\mathcal{G}_T$-measurable random variables.

Using the density $\alpha_t(\theta)$ allows us to calculate the $\mathbb{G}$-conditional expectations.

Proposition 1.5.2 *We assume that the (H)-hypothesis holds. Let $Y(T,\tau)$ be a random variable as above. Then we have*

1)

$$\mathbb{E}[Y(T,\tau)|\mathcal{G}_t]1\!\!1_{\{\tau>t\}} = \frac{\mathbb{E}[\int_t^\infty Y(T,u)\alpha_T(u)du|\mathcal{F}_t]}{\int_t^\infty \alpha_t(u)du}1\!\!1_{\{\tau>t\}}. \tag{1.22}$$

2)

$$\mathbb{E}[Y(T,\tau)|\mathcal{G}_t]1\!\!1_{\{\tau\leq t\}} = \mathbb{E}[Y(T,s)|\mathcal{F}_t]\big|_{s=\tau}1\!\!1_{\{\tau\leq t\}}. \tag{1.23}$$

Proof. 1) We know that $\mathbb{E}[Y(T,\tau)|\mathcal{G}_t]$ equals $\mathbb{E}[Y(T,\tau)|\mathcal{F}_t]/\mathbb{P}(\tau>t|\mathcal{F}_t)$ on $\{\tau>t\}$, which implies immediately (1.22) by the definition of $\alpha_t(\theta)$.

2) It suffices to prove for any $Y(T,\tau)$ of the form $Y(T,\tau)=Yg(\tau)$ where Y is an $\mathcal{F}_T$-measurable random variable and g is a Borel function. We need to verify that for any bounded $\mathcal{G}_t$-measurable random variable Z, we have

$$\mathbb{E}\big[Z1\!\!1_{\{\tau\leq t\}}\mathbb{E}[Yg(\tau)|\mathcal{G}_t]\big] = \mathbb{E}\big[Z\mathbb{E}[Y|\mathcal{F}_t]g(\tau)1\!\!1_{\{\tau\leq t\}}\big]. \tag{1.24}$$

By definition of conditional expectation, the left side of (1.24) equals to $\mathbb{E}[Z1\!\!1_{\{\tau\leq t\}}Yg(\tau)]$. On the other hand, (H)-hypothesis implies the independence between $\mathcal{F}_\infty$ and $\mathcal{G}_t$ conditioned on $\mathcal{F}_t$. So we have

$$\mathbb{E}[Y|\mathcal{F}_t]\mathbb{E}[Zg(\tau)1\!\!1_{\{\tau\leq t\}}|\mathcal{F}_t] = \mathbb{E}[YZg(\tau)1\!\!1_{\{\tau\leq t\}}|\mathcal{F}_t].$$

Therefore $\mathbb{E}\big[Z\mathbb{E}[Y|\mathcal{F}_t]g(\tau)1\!\!1_{\{\tau\leq t\}}\big] = \mathbb{E}[Z1\!\!1_{\{\tau\leq t\}}Yg(\tau)]$, which proves (1.24). □

Remark 1.5.3 We observe that the density $\alpha_t(\theta)$ is the key term for the calculation. The knowledge on the this density enables us to construct the conditional survival probability $\mathbb{P}(\tau>\theta|\mathcal{F}_t)$, which is our main tool to study $\mathbb{G}$-conditional expectations. Note that we have discussed the case where $\theta \geq t$ using the process G. Here we need also to study the case where $\theta < t$.

1.5.2 The general case with density

In this subsection, we suppose no longer that (H)-hypothesis holds. Instead, we introduce the following hypothesis that the conditional survival probability G_t^θ admits a density, which permits us to conduct explicit calculations. Indeed, it's convenient to work directly with the density $\alpha_t(\theta)$ instead of the intensity process as shown by the case with the (H)-hypothesis. The results of the previous subsection can be recovered in this case.

Hypothesis 1.5.4 For any $t, \theta \geq 0$, we assume that

1. the $\mathbb{F}$-martingale $(G_t^\theta = \mathbb{P}(\tau > \theta | \mathcal{F}_t), t \geq 0)$ admits a strictly positive density, that is, for any $\theta \geq 0$, there exists a strictly positive $\mathbb{F}$-adapted process $(\alpha_t(\theta), t \geq 0)$ such that
$$G_t^\theta = \int_\theta^\infty \alpha_t(u) du;$$
2. the process $(\alpha_t(\theta), t \geq 0)$ is an integrable $\mathbb{F}$-martingale on $[0, T]$.

The notion of the density $\alpha_t(\theta)$, which can be viewed as some **martingale density**, plays the crucial role in our further discussions. Furthermore, it provides a general method which adapts without any difficulty in the multi-credits to study the successive defaults.

By introducing this density, we can calculate the $\mathbb{G}$-conditional expectations, even on $\{\tau \leq t\}$, where the explicit form contains the quotient of two densities. The general formula is given in Theorem 1.5.5. Moreover, we obtain the compensator process explicitly in Theorem 1.5.7. The martingale density is an efficient tool to study the case after the default. In fact, by comparing (1.25) and (1.26), we observe some similitude between the two formulae, which shows that we can study the cases before and after the default in the same framework we introduce.

Theorem 1.5.5 *Let $Y(T, \tau)$ be an integrable $\mathcal{G}_T$-measurable random variable. Then for any $0 \leq t \leq T$,*

1)

$$\mathbb{E}[Y(T,\tau)|\mathcal{G}_t]1\!\!1_{\{\tau>t\}} = \frac{\mathbb{E}\big[\int_t^\infty Y(T,u)\alpha_T(u)du|\mathcal{F}_t\big]}{\int_t^\infty \alpha_t(u)du}1\!\!1_{\{\tau>t\}} = \frac{\mathbb{E}\big[\int_t^\infty Y(T,u)\alpha_T(u)du|\mathcal{F}_t\big]}{G_t}1\!\!1_{\{\tau>t\}}. \tag{1.25}$$

2) Recall that we have supposed that $\alpha_t(\theta) > 0$. Then

$$\mathbb{E}[Y(T,\tau)|\mathcal{G}_t]1\!\!1_{\{\tau\leq t\}} = \mathbb{E}\Big[Y(T,s)\frac{\alpha_T(s)}{\alpha_t(s)}\big|\mathcal{F}_t\Big]\Big|_{s=\tau}1\!\!1_{\{\tau\leq t\}} = \frac{\mathbb{E}\big[Y(T,s)\alpha_T(s)\big|\mathcal{F}_t\big]}{\alpha_t(s)}\Big|_{s=\tau}1\!\!1_{\{\tau\leq t\}}. \tag{1.26}$$

The density $(\alpha_t(\theta), t \geq 0)$ is an $\mathbb{F}$-martingale and $\alpha_t(\theta) = 0$ implies that $\alpha_T(\theta) = 0$ for any $T > t$.

Proof. We here only give the proof of 2). In fact, it suffices to verify for $Y(T,\tau)$ of the form $Y(T,\tau) = Yg(\tau \wedge T)$ where Y is an integrable $\mathcal{F}_T$-measurable random variable and g is a bounded Borel function. The proof is similar to that of Proposition 1.5.2. We shall verify, for any bounded $\mathcal{G}_t$-measurable random variable Z, that

$$\mathbb{E}\big[Z\mathbb{E}[Yg(\tau)|\mathcal{G}_t]1\!\!1_{\{\tau\leq t\}}\big] = \mathbb{E}\Big[Z\mathbb{E}[Y\frac{\alpha_T(s)}{\alpha_t(s)}|\mathcal{F}_t]\Big|_{s=\tau} g(\tau)1\!\!1_{\{\tau\leq t\}}\Big].$$

Since $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{D}_t$, there exists a bounded $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}_t$-measurable function F on $\mathbb{R}_+ \times \Omega$ such that $Z = F(\tau \wedge t, \omega)$,

$$\mathbb{E}\big[Z\mathbb{E}[Yg(\tau)|\mathcal{G}_t]1\!\!1_{\{\tau\leq t\}}\big] = \mathbb{E}\big[\mathbb{E}[ZYg(\tau)1\!\!1_{\{\tau\leq t\}}|\mathcal{F}_T]\big] = \mathbb{E}\Big[Y\int_0^t F(s,\omega)g(s)\alpha_T(s)ds\Big].$$

On the other hand,

$$\mathbb{E}\Big[Z\mathbb{E}\big[Y\frac{\alpha_T(s)}{\alpha_t(s)}|\mathcal{F}_t\big]\Big|_{s=\tau} g(\tau)1\!\!1_{\{\tau\leq t\}}\Big] = \mathbb{E}\Big[\int_0^t F(s,\omega)\mathbb{E}\big[Yg(s)\frac{\alpha_T(s)}{\alpha_t(s)}|\mathcal{F}_t\big]\alpha_t(s)ds\Big]$$
$$= \mathbb{E}\big[\int_0^t F(s,\omega)\mathbb{E}[Yg(s)\alpha_T(s)|\mathcal{F}_t]ds\big] = \mathbb{E}\big[Y\int_0^t F(s,\omega)g(s)\alpha_T(s)ds\big].$$

□

Remark 1.5.6 In the first equality of (1.26), we deal with $\alpha_T(\theta)/\alpha_t(\theta)$, which is the quotient of martingale densities. With this quantity, we can calculate the $\mathbb{G}$-expectation for a r.v. Y of interest. In the second equality of (1.26), we first take conditional expectation and we deal with the quotient of another $\mathbb{F}$-martingale with respect to $\alpha_t(\theta)$. The second way is less convenient for calculation. However, it can be extended to the case without density.

Since $(\alpha_t(\theta), t \geq 0)$ is a uniformly integrable martingale, we consider the change of probability defined by $d\mathbb{Q}^s = \alpha_T(s)d\mathbb{P}$ on $\mathcal{F}_T$, then (1.26) implies that

$$\mathbb{E}[Y(T,\tau)|\mathcal{G}_t]1\!\!1_{\{\tau\leq t\}} = \mathbb{E}_{\mathbb{Q}^s}\big[Y(T,s)|\mathcal{F}_t\big]\big|_{s=\tau}1\!\!1_{\{\tau\leq t\}}.$$

This is similar with the discussion on the process G in the previous chapter with the change of probability $\mathbb{Q}^L$, where we interpret the relationship between the (H)-hypothesis. Note that when the (H)-hypothesis holds, we have (1.23), which is of the same form as the above formula under the initial probability. So there is no need to change the probability measure under the (H)-hypothesis. In addition, we can rewritten (1.25) as follows.

$$\begin{aligned}\mathbb{E}[Y(T,\tau)|\mathcal{G}_t]1\!\!1_{\{\tau>t\}} &= \frac{\int_t^\infty \mathbb{E}\big[Y(T,u)\alpha_T(u)|\mathcal{F}_t\big]du}{\int_t^\infty \alpha_t(u)du}1\!\!1_{\{\tau>t\}} \\ &= \frac{\int_t^\infty \mathbb{E}_{\mathbb{Q}^s}\big[Y(T,u)|\mathcal{F}_t\big]\alpha_t(u)du}{\int_t^\infty \alpha_t(u)du}1\!\!1_{\{\tau>t\}}.\end{aligned}$$

The following result deals with the calculation of the compensator process Λ of τ. By Theorem 1.3.4, we know that we need to find the Doob-Meyer decomposition of the process G. We now treat the case with density.

Theorem 1.5.7 *Assume that* $(\mathbb{F},\mathbb{G},\tau)$ *satisfy MA.*
We suppose that Hypothesis 1.5.4 holds. Then the $\mathbb{G}$*-compensator process* Λ *of* τ *is given by*

$$d\Lambda_t = 1\!\!1_{]0,\tau]}(t)\frac{\alpha_t(t)}{G_t}dt = 1\!\!1_{]0,\tau]}(t)\frac{\alpha_t(t)}{\int_t^\infty \alpha_t(u)du}dt. \tag{1.27}$$

Proof. We first notice that $(A_t = \int_0^t \alpha_v(v)dv, t \geq 0)$ is the increasing $\mathbb{F}$-predictable process of the Doob-Meyer decomposition of G. Or in other words, $(G_t+\int_0^t \alpha_v(v)dv, t \geq 0)$ is an $\mathbb{F}$-martingale. To prove this, it suffices to verify that

$$\begin{aligned}\mathbb{E}\big[G_T - G_t + \int_t^T \alpha_v(v)dv|\mathcal{F}_t\big] &= \mathbb{E}\big[\int_T^\infty \alpha_T(v)dv - \int_t^\infty \alpha_t(v)dv + \int_t^T \alpha_v(v)dv|\mathcal{F}_t\big] \\ &= \mathbb{E}\big[-\int_t^T \alpha_t(v)dv + \int_t^T \alpha_v(v)dv|\mathcal{F}_t\big] = 0.\end{aligned}$$

The last equality is due to the $\mathbb{F}$-martingale property of $(\alpha_t(\theta), t \geq 0)$. In addition, since G is continuous, then Theorem 1.3.4 implies immediately (1.27). □

Remark 1.5.8 1. Notice that

$$-\partial_\theta \ln \mathbb{P}(\tau > \theta|\mathcal{F}_t)\big|_{\theta=t} = -\partial_\theta\Big(\ln\int_\theta^\infty \alpha_t(u)du\Big)\Big|_{\theta=t} = \frac{\alpha_t(t)}{\int_t^\infty \alpha_t(u)du},$$

which is of form of a “real intensity” if we adopted the exponential notation.

2. Since both the Minimal Assumption and Hypothesis 1.5.4 can be generalized to the multi-credits case, we shall see in the next chapter that, thanks to the notion of the martingale density, the above theorem can be extended easily to study the successive defaults.

We now revisit the multiplicative decomposition of the process G. Notice that the process Λ is continuous here. In addition, using the forward intensity, this decomposition implies a HJM type result similar with (1.17) without the hypothesis that $\mathbb{F}$ is generated by a Brownian motion.

Corollary 1.5.9 *We use the notations of Proposition 1.3.6. Under Hypothesis 1.5.4, the multiplicative decomposition of G is given by*

$$G_t = \mathcal{E}(\widetilde{M}_t^{\widetilde{\Gamma}}) \exp\Big(- \int_0^t \gamma(s,s) ds \Big) \tag{1.28}$$

where $\gamma(t,T)$ is the $\mathbb{F}$-forward intensity of τ.

Proof. By proposition 1.3.6 and Theorem 1.5.7, we know that

$$G_t = \mathcal{E}(\widetilde{M}_t^{\widetilde{\Gamma}}) \exp\Big(- \int_0^t \frac{\alpha_s(s)}{G_s} ds \Big).$$

On the other hand, $\gamma(t,T) = -\partial_T \ln G_t^T = \frac{\alpha_t(T)}{G_t^T}$, which implies directly (1.28). □

1.5.3 The general case without density

We have supposed until now the hypothesis that the density $\alpha_t(\theta)$ exists and is strictly positive. Note that the density $\alpha_t(\theta)$ is related to the conditional survival probability. In this subsection, we show that by adopting the point of view of some "random measure" to represent the conditional expectation, the above hypotheses are not necessary.

Definition 1.5.10 We define an *$\mathcal{F}_t$-measure* q_t to be the continuous linear application from $M_b(\mathbb{R}_+)$ to $L^1(\mathcal{F}_t)$ such that

$$q_t(f) = \mathbb{E}[f(\tau)|\mathcal{F}_t] \tag{1.29}$$

where $M_b(\mathbb{R}_+)$ is the set of all bounded Borel functions on $\mathbb{R}_+$.

Remark 1.5.11 1) In the case with density, we have

$$q_t(f) = \int_{\mathbb{R}_+} f(\theta)\alpha_t(\theta) d\theta.$$

2) By definition, we have $\mathbb{E}[q_T(f)|\mathcal{F}_t] = q_t(f)$. For simplicity, we denote by $\mathbb{E}[q_T|\mathcal{F}_t] = q_t$. Hence, $(q_t, t \geq 0)$ is a measure-valued $\mathbb{F}$-martingale in the sense that for any $f \in M_b(\mathbb{R}_+)$, $(q_t(f), t \geq 0)$ is an $\mathbb{F}$-martingale.

When calculating the conditional expectations on $\{\tau \leq t\}$ in Proposition 1.5.5, the key term is $\alpha_T(s)/\alpha_t(s)$ which is the ratio of two densities. This kind of ratio appears naturally in the comparison of two measures on $\mathbb{R}_+$ having densities. In the general case, we can also compare two measures by the Radon-Nikodym theorem.

By Definition 1.5.10, we have defined a general measure q_t without density. If we draw some analogy with Proposition 1.5.5, we are interested in the comparison of measures dq_T/dq_t instead of $\alpha_T(s)/\alpha_t(s)$. However, we here encounter technical difficulty since q_T is a $\mathcal{F}_T$-measure and q_t is a $\mathcal{F}_t$-measure. We hence propose an alternative method which consists of comparing two $\mathcal{F}_t$-measures (analogous with $\mathbb{E}[Y(T,s)\alpha_T(s)|\mathcal{F}_t]$ and $\alpha_t(s)$ in (1.26)).

Similarly with Definition 1.5.10, we now define q_t^Y, for any integrable $\mathcal{G}$-measurable random variable Y, to be a continuous linear application from $M_b(\mathbb{R}_+)$ to $L^1(\mathcal{F}_t)$ such that

$$q_t^Y(f) = \mathbb{E}[Yf(\tau)|\mathcal{F}_t]. \tag{1.30}$$

The explicit form of q^Y depends on the measurability of Y. Under certain condition, it can be determined by the martingale measure $(q_t, t \geq 0)$. For example, if Y is $\mathcal{F}_T$-measurable, we have

$$q_t^Y(f) = \mathbb{E}[Yq_T(f)|\mathcal{F}_t].$$

In addition, in the case with density, we have $q_t^Y(f) = \mathbb{E}\big[Y\int_{\mathbb{R}_+} f(\theta)\alpha_T(\theta)d\theta|\mathcal{F}_t\big]$ if Y is $\mathcal{F}_T$-measurable.

As suggested by (1.26), we would like to prove a result of the form

$$\mathbb{E}[Yg(\tau)|\mathcal{G}_t]1\!\!1_{\{\tau\leq t\}} = \frac{dq_t^Y}{dq_t}(\tau)g(\tau)1\!\!1_{\{\tau\leq t\}}$$

and in the case with density, we shall have

$$\frac{dq_t^Y}{dq_t}(s) = \mathbb{E}\Big[Y\frac{\alpha_T(s)}{\alpha_t(s)}\big|\mathcal{F}_t\Big].$$

There are two major difficulties: 1) what is the analogy of the absolute continuity of an $\mathcal{F}_t$-measure with respect to another; 2) how to define the Radon-Nikodym derivative for dq_t^Y/dq_t.

To interpret dq_t^Y/dq_t as the classical Radon-Nikodym derivative, we introduce some classical measures on the product space $(\mathbb{R}_+\times\Omega, \mathcal{B}(\mathbb{R}_+)\otimes\mathcal{F}_t)$ for any $F\in M_b(\mathbb{R}_+\times\Omega, \mathcal{B}(\mathbb{R}_+)\otimes\mathcal{F}_t)$:

$$\mu_t(F(s,\omega)) = \mathbb{E}\big[F(\tau(\omega),\omega)\big] \quad \text{and} \quad \mu_t^Y(F(s,\omega)) = \mathbb{E}\big[YF(\tau(\omega),\omega)\big],$$

which can be viewed as the conditional expectation with respect to $\sigma(\tau)\vee\mathcal{F}_t$.

If Y is $\mathcal{F}_T$-measurable, then μ_t and μ_t^Y can be determined by the martingale measure $(q_t, t \geq 0)$ under some conditions. For example, when $F(s,\omega) = f(s)Z(\omega)$ with f a Borel function on $\mathbb{R}_+$ and Z an $\mathcal{F}_t$-measurable r.v., we have

$$\mu_t(F) = \mathbb{E}[Zq_t(f)] \quad \text{and} \quad \mu_t^Y(F) = \mathbb{E}[YZq_T(f)].$$

In general, for any function in $M_b(\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}_t)$, we only need to consider linear combinations of functions of the above form.

Lemma 1.5.12 *For any $t \geq 0$, we have $\mu_t^Y \ll \mu_t$.*

Proof. For any $F \geq 0$ such that $\mathbb{E}[F(\tau,\omega)] = 0$, we have $F(\tau,\omega) = 0$ a.s.. Then $\mathbb{E}[YF(\tau,\omega)] = 0$. □

Proposition 1.5.13 *Let Y be an integrable $\mathcal{G}$-measurable random variable and let g be a bounded Borel function. Then for any $t \geq 0$,*

$$\mathbb{E}[Yg(\tau)|\mathcal{G}_t]1\!\!1_{\{\tau\leq t\}} = \frac{d\mu_t^Y}{d\mu_t}(\tau,\omega)g(\tau)1\!\!1_{\{\tau\leq t\}}.$$

Proof. For any bounded $\mathcal{G}_t$-measurable random variable $Z = F(\tau \wedge t, \omega)$, where F is a bounded $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}_t$-measurable on $\mathbb{R}_+ \times \Omega$, we have

$$\begin{aligned}
&\mathbb{E}[ZYg(\tau)1\!\!1_{\{\tau\leq t\}}] = \int_{[0,t]\times\Omega} F(s,\omega)g(s)d\mu_t^Y(s,\omega) \\
&= \int_{[0,t]\times\Omega} \frac{d\mu_t^Y}{d\mu_t}(s,\omega)F(s,\omega)g(s)d\mu_t(s,\omega) = \mathbb{E}\Big[F(\tau,\omega)\frac{d\mu_t^Y}{d\mu_t}(\tau,\omega)g(\tau)1\!\!1_{\{\tau\leq t\}}\Big] \\
&= \mathbb{E}\Big[Z\frac{d\mu_t^Y}{d\mu_t}(\tau,\omega)g(\tau)1\!\!1_{\{\tau\leq t\}}\Big].
\end{aligned}$$

□

Remark 1.5.14 In the case with density, $d\mu_t^Y/d\mu_t$ can be calculated explicitly as

$$\frac{d\mu_t^Y}{d\mu_t} = \frac{\partial_s\mathbb{E}[Y1\!\!1_{\{\tau>s\}}|\mathcal{F}_t]}{\partial_s\mathbb{E}[1\!\!1_{\{\tau>s\}}|\mathcal{F}_t]}.$$

In this case, μ_t is absolutely continuous with respect to the canonical measure on $(\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}_t)$ and the function $\mathbb{E}[1\!\!1_{\{\tau<s\}}|\mathcal{F}_t]$ is absolutely continuous on s. Moreover, the result of the Proposition 1.5.13 can be written as

$$\mathbb{E}[Yg(\tau)|\mathcal{G}_t]1\!\!1_{\{\tau\leq t\}} = \frac{\partial_s\mathbb{E}[Y1\!\!1_{\{\tau>s\}}|\mathcal{F}_t]}{\partial_s\mathbb{E}[1\!\!1_{\{\tau>s\}}|\mathcal{F}_t]}\Big|_{\{s=\tau\}}g(\tau)1\!\!1_{\{\tau\leq t\}}. \tag{1.31}$$

Chapter 2

The multi-defaults case

This chapter is motivated by some practical concern of the practitioners on the market: when managing a portfolio which contains several defaultable credits, one observes essentially the occurrence of the first default. Hence, to study the successive defaults in the intensity approach, can one suppose that the further defaults preserve the probability law of the same family (usually the exponential family in practice) with nevertheless parameters depending on the market data of the observation time?

We develop a simple deterministic model of two-credits to study this problem and we show that this kind of properties hold only in very special cases and the associated copula function depends on the dynamic of the marginal distributions. We deduce the distribution of the second default time and observe that the calculations become complicated and the result is not clear to interpret. Hence, we find the intuition of the market is inappropriate to model the multi-credits case.

In the second section, we study two default times in the general framework which we proposed in the previous chapter. We deduce the compensator process of the second default time with respect to the first one and we show that the result can be extended without difficulty to the successive defaults. Hence, instead of the classical procedure where we treat first the marginal distributions of each credit and then their joint distribution, we propose an original method which concentrates on the successive defaults.

2.1 An illustrative model

2.1.1 Model setup

In this subsection, we present a very simple model from the practical point of view. On the market, the practitioners adopt more intensity models than structural models because the intensity models fit easily to the daily data of CDS spreads. A simplified but largely used version of the intensity model is the exponential model where the default or survival probability of one credit is calculated by an exponential distribution with parameters being calibrated from the CDS market. The computation is repeated each day with daily CDS data.

Using this procedure, the practitioners adopt some time stationary property. This idea is extended to the portfolio case. That is, when we consider a portfolio containing several credits, we suppose that each credit satisfies the exponential distribution hypothesis. What we need to specify is the observable information. In the single credit case, we observe the default-or-not event of the credit concerned. When the default occurs, there will be no need to calculate the conditional survival probability of course. However, in the multi-credit case, after the first default, we shall calculate the survival probabilities of the other credits conditionned on this event. Before the first default, each credit satisfies the exponential hypothesis. For this practical reason, we introduce an hypothesis on the joint law given by (2.1). Moreover, in practice, it is often supposed that the surviving credits still satisfy this condition after the first default occurs, once the parameters having been adjusted to the "contagious jump" phenomenon. We shall discuss this argument in subsection 2.1.3.

In the following, we consider two credits and we suppose that the available information is whether the first default occurs or not. Denote by τ_1 and τ_2 the default times of each credit and let $\tau = \min(\tau_1, \tau_2)$. Each default time is supposed to follow an exponential type distribution of parameter $\mu^i(t)$ before the first default occurs. That is, for any $T > t$, we suppose

$$\mathbb{P}(\tau_i > T \mid \tau > t) = e^{-\mu^i(t)\cdot(T-t)}, \quad (i = 1, 2). \tag{2.1}$$

Notice that $\mu^i(t)$ is a deterministic function. In fact, before the first default, we are in the deterministic context and all the conditional expectations can be calculated explicitly with conditional probabilities. By letting $t = 0$, we have

$$\mathbb{P}(\tau_i > T) = e^{-\mu^i(0)T}. \tag{2.2}$$

At the initial time, each credit follows the exponential law with intensity $\mu^i(0)$. At time t, $\mu^i(t)$ is renewed with observed information. When τ_1 and τ_2 are independent, it's easy to calculate the joint probability

$$\mathbb{P}(\tau_1 > t_1, \tau_2 > t_2) = \mathbb{P}(\tau_1 > t_1)\mathbb{P}(\tau_2 > t_2) = e^{-\mu^1(0)t_1 - \mu^2(0)t_2}.$$

Then we obtain immediately $\mu^i(t) = \mu^i(0)$, which means $\mu^i(t)$ remains constant in the independent case.

Hypothesis (2.1) shows the stationary property of the individual default distribution. Notice that $\mu^i(t) = -\frac{1}{T-t}\ln \mathbb{P}(\tau_i > T|\tau > t)$ can be viewed as the *implied hazard rate* (cf. [73]) without interest rate and recovery rate. In this model, (2.1) implies that this forward rate does not depend on T.

2.1.2 The joint distribution

In this subsection, we show that the joint probability distribution can be determined explicitly in this model. To show the relationship between the joint and the marginal distributions, we write the joint probability $\mathbb{P}(\tau_1 > t_1, \tau_2 > t_2)$ as the product of two marginal probabilities and a function $\rho(t_1, t_2)$ which represents the correlation between them, i.e.,

$$\mathbb{P}(\tau_1 > t_1, \tau_2 > t_2) = \mathbb{P}(\tau_1 > t_1)\mathbb{P}(\tau_2 > t_2)\rho(t_1, t_2). \tag{2.3}$$

Note that contrary to the linear correlation parameter which takes value in $[0, 1]$, the function $\rho(t_1, t_2)$ can take any strictly positive real value. In addition, we have $\rho(0, t) = \rho(t, 0) = 1$ for any $t \geq 0$. In particular, for any $t_1, t_2 > 0$, if $\rho(t_1, t_2) = 1$, then there is independence.

In fact, (2.3) defines a unique copula function for any t_1 and t_2 by the Sklar's theorem. Let $\widetilde{C}(u, v)$ with $(u, v) \in [0, 1]^2$ be the survival copula function such that the joint probability defined by (2.3) can be written as

$$\widetilde{C}(\mathbb{P}(\tau_1 > t_1), \mathbb{P}(\tau_2 > t_2)) = \mathbb{P}(\tau_1 > t_1)\mathbb{P}(\tau_2 > t_2)\rho(t_1, t_2).$$

By letting $u = \mathbb{P}(\tau_1 > t_1)$ and $v = \mathbb{P}(\tau_2 > t_2)$ and by using (2.2), we get

$$\widetilde{C}(u, v) = \begin{cases} uv\rho\Big(\frac{\ln u}{\mu^1(0)}, \frac{\ln v}{\mu^2(0)}\Big), & \text{if} \quad u, v > 0; \\ 0, & \text{if} \quad u = 0 \quad \text{or} \quad v = 0. \end{cases}$$

Hence $\widetilde{C}$ is a special copula function which depends on the initial values of $\mu^i(t)$ and the function ρ. We are interested in the form of the function ρ since it implies directly the joint survival probability function.

In the standard copula model, the joint probability function depends on μ^i through the marginal probability which is a uniform variable $\mathbb{P}(\tau_i > t_i) = e^{-\mu^i(0)t_i}$, that is, only the initial values $\mu^1(0)$ and $\mu^2(0)$. In the following, we show that because of the function ρ which can be deduced explicitly by hypothesis (2.1), the joint probability in this case depends not only the initial values, but on the functions $\mu^1(t)$ and $\mu^2(t)$.

Proposition 2.1.1 *If $\rho(t_1, t_2) \in C^{1,1}$, then the joint survival probability is given by*

$$\mathbb{P}(\tau_1 > t_1, \tau_2 > t_2) = \exp\Big(-\int_0^{t_1} \mu^1(s \wedge t_2)ds - \int_0^{t_2} \mu^2(s \wedge t_1)ds\Big). \tag{2.4}$$

Proof. From (2.1), we have $\mathbb{P}(\tau_1 > T, \tau_2 > t) = \mathbb{P}(\tau_1 > t, \tau_2 > t)e^{-\mu^1(t)(T-t)}$ for any $T > t$. Combining the definition (2.3), we obtain

$$\frac{\rho(T, t)}{\rho(t, t)} = e^{-(\mu^1(t)-\mu^1(0))(T-t)}.$$

Then $\partial_T \ln \rho(T,t) = -(\mu^1(t) - \mu^1(0))$. By symmetry, we also have $\partial_T \ln \rho(t,T) = -(\mu^2(t)-\mu^2(0))$. Taking the sum, we get $\frac{d}{dt}\ln \rho(t,t) = -(\mu^1(t)-\mu^1(0))-(\mu^2(t)-\mu^2(0))$ and then

$$\ln \rho(t,t) = -\int_0^t (\mu^1(s) - \mu^1(0)) + (\mu^2(s) - \mu^2(0))ds.$$

Therefore, $\ln \rho(T,t) = \mu^1(0)T+\mu^2(0)t-\mu^1(t)(T-t)-\int_0^t(\mu^1(s)+\mu^2(s))ds$ and $\ln \rho(t,T)$ is obtained by symmetry. Then for any $t_1, t_2 \geq 0$,

$$\ln \rho(t_1,t_2) = \mu^1(0)t_1 + \mu^2(0)t_2 - \int_0^{t_1} \mu^1(s \wedge t_2)ds - \int_0^{t_2} \mu^2(s \wedge t_1)ds,$$

which implies (2.4). □

Remark 2.1.2 The above proposition shows that the joint probability function $\mathbb{P}(\tau_1 > t_1, \tau_2 > t_2)$ depends on the marginal distributions through $(\mu^1(t), t \leq t_1)$ and $(\mu^2(t), t \leq t_2)$, which means that it depends on all the marginal dynamics.

Proposition 2.1.3 *If $\rho(t_1,t_2) \in C^2$ and if $\mu^1(t), \mu^2(t) \in C^1$, then*

$$\mu^i(t) = \mu^i(0) - \int_0^t \varphi(s)ds \tag{2.5}$$

where

$$\varphi(t) = \frac{\partial^2}{\partial t_1 \partial t_2}\Big|_{t_1=t_2=t} \ln \rho(t_1,t_2).$$

In addition, we have

$$\mathbb{P}(\tau_1 > t_1, \tau_2 > t_2) = \exp\Big(- \mu^1(0)t_1 - \mu^2(0)t_2 + \int_0^{t_1 \wedge t_2} \varphi(s)(t_1+t_2-2s)ds\Big). \tag{2.6}$$

Proof. Notice that when $t_1 \leq t_2$, $\partial^2_{t_1,t_2} \ln \rho(t_1,t_2) = -(\mu^2)'(t_1)$ and when $t_1 \geq t_2$, $\partial^2_{t_1,t_2} \ln \rho(t_1,t_2) = -(\mu^1)'(t_2)$. Then

$$\partial^2_{1,2}|_{\{t_1=t_2=t\}} \ln \rho(t_1,t_2) = -(\mu^1)'(t) = -(\mu^2)'(t).$$

By the definition of $\varphi(t)$, we have $(\mu^1)'(t) = (\mu^2)'(t) = -\varphi(t)$, which implies immediately (2.5). By replacing $\mu^i(t)$ in equation (2.4) with the integral form (2.5) and taking integration by part, we get (2.6). □

Remark 2.1.4 1) In the above proposition, (2.5) is rather astonishing at the first sight because it means that $\mu^1(t)$ and $\mu^2(t)$ are identical apart from their initial values. The point lies in the stationary property implied by (2.1) and the fact that the information flow is symmetric for the two credits before the first default.

2) We obtain, as a direct consequence of (2.6), the explicit form of $\rho(t_1,t_2)$ given by $\rho(t_1,t_2) = \exp\big(\int_0^{t_1\wedge t_2}\varphi(s)(t_1+t_2-2s)ds\big)$. Therefore, the function φ plays an important role in determining the correlation structure of default times.

Mathematical criteria are required to well define (2.6).

i) the survival probability should be decreasing with respect to time, which implies

$$\frac{\partial}{\partial t_i}\ln\mathbb{P}(\tau_1>t_1,\tau_2>t_2)\le 0,\quad \forall t_1,t_2\in[0,T];$$

ii) the probability density function should be positive, which implies

$$\frac{\partial^2}{\partial t_1\partial t_2}\mathbb{P}(\tau_1>t_1,\tau_2>t_2)\ge 0.$$

Therefore, the function φ satisfies the following conditions:

$$\begin{cases}-\mu^1(0)+\int_0^{t_2}\varphi(s\wedge t_1)ds\le 0\\ -\mu^2(0)+\int_0^{t_1}\varphi(s\wedge t_2)ds\le 0\end{cases}\tag{2.7}$$

and

$$\varphi(t_1\wedge t_2)\ge -\left(\mu^1(0)-\int_0^{t_2}\varphi(s\wedge t_1)ds\right)\left(\mu^2(0)-\int_0^{t_1}\varphi(s\wedge t_2)ds\right).\tag{2.8}$$

We notice that condition (2.8) is always satisfied when $\varphi\ge 0$. When $\varphi=0$, there is independence since $\mathbb{P}(\tau_1>t_1,\tau_2>t_2)=\mathbb{P}(\tau_1>t_1)\mathbb{P}(\tau_2>t_2)$. In addition, the function $\mu^i(t)$ remains constant as the initial value.

Remark 2.1.5 If i) and ii) are satisfied, then the right-hand side of (2.6) defines a joint probability distribution on $\mathbb{R}^2$. Denote by $G(t_1,t_2)=\mathbb{P}(\tau_1>t_1,\tau_2>t_2)$. In fact, it's not difficult to verify that $G(0,0)=1$ and $\lim_{t_1\to+\infty}G(t_1,t_2)=\lim_{t_2\to+\infty}G(t_1,t_2)=0$. Since by i), $G(t_1,t_2)$ is decreasing with respect to t_1 and to t_2, $\lim_{t_1\to+\infty,t_2\to+\infty}G(t_1,t_2)=0$. Moreover, for any $0\le x_1\le x_2$ and $0\le y_1\le y_2$, we have

$$\begin{aligned}\mathbb{P}(x_1<\tau_1\le x_2,y_1<\tau_2\le y_2)&=G(x_1,y_1)-G(x_2,y_1)-G(x_1,y_2)+G(x_2,y_2)\\&=\int_{x_1}^{x_2}\int_{y_1}^{y_2}\partial^2_{1,2}G(x,y)dxdy.\end{aligned}$$

2.1.3 First default and contagious jumps

In this subsection, we study the first default and its impact. First, we can easily obtain the probability distribution of the first default time by (2.4) and (2.6),

$$\begin{aligned}\mathbb{P}(\tau>t)&=\exp\Big(-\int_0^t\mu^1(s)+\mu^2(s)ds\Big)\\&=\exp\Big(-\big(\mu^1(0)+\mu^2(0)\big)t+2\int_0^t\varphi(s)(t-s)ds\Big).\end{aligned}\tag{2.9}$$

Note that the first default time also follows an exponential law and $\mu^1(t) + \mu^2(t)$ can be viewed as some intensity parameter.

When the first default occurs, the following result of Jeanblanc [52] (see also [79]) enables us to calculate conditional probability of the surviving credit.

Proposition 2.1.6 *Let $\mathcal{D}_t = \mathcal{D}_t^1 \vee \mathcal{D}_t^2$ where $\mathcal{D}_t^i = \sigma(1\!\!1_{\{\tau_i \leq s\}}, s \leq t)$ $(i = 1, 2)$. Denote by $G(x, y) = \mathbb{P}(\tau_1 > x, \tau_2 > y)$. If $G(x, y) \in C^{1,1}$ on $\mathbb{R}^2$, then*

$$1\!\!1_{\{\tau \leq t\}} \mathbb{E}[1\!\!1_{\{\tau_1 > T\}} \mid \mathcal{D}_t] = 1\!\!1_{\{\tau_2 \leq t, \tau_1 > t\}} \frac{\partial_2 G(T, \tau)}{\partial_2 G(t, \tau)}, \tag{2.10}$$

Proof. First, we have for any $s \leq t$ that

$$\mathbb{E}[1\!\!1_{\{\tau_2 > s\}} 1\!\!1_{\{\tau \leq t\}} \mathbb{E}[1\!\!1_{\{\tau_1 > T\}} \mid \mathcal{D}_t]] = \mathbb{P}(\tau_1 > T, s < \tau_2 \leq t) = G(T, s) - G(T, t). \tag{2.11}$$

In addition, since $\partial^2_{x,y} G(x, y)$ is the probability density function of τ_1 and τ_2, we have

$$\begin{aligned}
\mathbb{E}\left[1\!\!1_{\{\tau_2 > s\}} 1\!\!1_{\{\tau_2 \leq t, \tau_1 > t\}} \frac{\partial_2 G(T, \tau)}{\partial_2 G(t, \tau)}\right] &= \int_s^t \int_t^\infty \frac{\partial_2 G(T, y)}{\partial_2 G(t, y)} \frac{\partial^2}{\partial x \partial y} G(x, y) dx dy \\
&= -\int_s^t \partial_2 G(T, y) dy = G(T, s) - G(T, t).
\end{aligned}$$

Then by the definition of conditional expectation, we get (2.10). □

Combining the above proposition and (2.6), we get the marginal conditional probabilities.

Proposition 2.1.7 *For any $t \geq 0$ and $T > t$, the conditional survival probability of the credit i is given by*

$$\begin{aligned}
&\mathbb{E}[1\!\!1_{\{\tau_i > T\}} \mid \mathcal{D}_t] = 1\!\!1_{\{\tau > t\}} \exp\left(-\big(\mu^i(0) - \int_0^t \varphi(s) ds\big)(T - t)\right) \\
&+ 1\!\!1_{\{\tau_i > t, \tau_j \leq t\}} \exp\left(-\big(\mu^i(0) - \int_0^\tau \varphi(s) ds\big)(T - t)\right) \cdot \frac{\mu^j(0) - \varphi(\tau)(T - \tau) - \int_0^\tau \varphi(s) ds}{\mu^j(0) - \varphi(\tau)(t - \tau) - \int_0^\tau \varphi(s) ds}.
\end{aligned} \tag{2.12}$$

Proof. We write $\mathbb{E}[1\!\!1_{\{\tau_i > T\}} | \mathcal{D}_t] = 1\!\!1_{\{\tau > t\}} \mathbb{E}[1\!\!1_{\{\tau_i > T\}} | \mathcal{D}_t] + 1\!\!1_{\{\tau_i > t, \tau_j \leq t\}} \mathbb{E}[1\!\!1_{\{\tau_i > T\}} | \mathcal{D}_t]$. To calculate the first term, it suffices to recall $\mu^i(t)$ given by (2.5). For the second term, we calculate the conditional expectation by (2.11) in the above proposition and the joint survival probability function given by (2.6). □

Remark 2.1.8 1. Notice that $\mathbb{E}[1\!\!1_{\{\tau_i > T\}} | \mathcal{D}_t]$ can be viewed as a defaultable zero coupon price. The two terms at the right-hand side of (2.12) represents respectively the price before and after the default of the other credit. By comparing the two terms, we observe the so-called "contagious default" phenomenon since the survival probability of the i^{th}-credit has a jump downward at τ given the default of the other credit if $\varphi > 0$.

This phenomenon has been discussed by Jarrow and Yu [51] where the authors supposed that the intensity process have a positive jump when the other credit defaults.

2. After the first default, the surviving credit satisfies no longer the exponential stationary property by the form of the conditional probability in (2.12). So in general, we can not expect to proceed in a recursive way since the basic hypothesis is no longer valid.

We are now interested in the properties of the second default time with respect to the first one. We denote by

$$\sigma = \max(\tau_1, \tau_2).$$

Let us define the filtration $\mathbb{D}^\tau = (\mathcal{D}_t^\tau)_{t\geq 0}$ associated with the first default time τ where $\mathcal{D}_t^\tau = \sigma(1\!\!1_{\{\tau \leq s\}}, s \leq t)$.

Proposition 2.1.9 *For any $t \geq 0$ and $T \geq t$, we have*

$$\begin{aligned}
&\mathbb{E}[1\!\!1_{\{\sigma > T\}} | \mathcal{D}_t^\tau] \\
&= 1\!\!1_{\{\tau > t\}} \Big[\exp\big(-\mu^1(t)(T-t)\big) + \exp\big(-\mu^2(t)(T-t)\big) - \exp\Big(-\int_t^T \big(\mu^1(s)+\mu^2(s)\big)ds\Big)\Big] \\
&+ 1\!\!1_{\{\tau \leq t\}} \Bigg[\exp\big(-\mu^1(\tau)(T-\tau)\big) \frac{\mu^2(0) - \int_0^T \varphi(\tau\wedge s)ds}{\mu^1(\tau)+\mu^2(\tau)} \\
&\qquad + \exp\big((-\mu^2(\tau)(T-\tau)\big) \frac{\mu^1(0) - \int_0^T \varphi(\tau\wedge s)ds}{\mu^1(\tau)+\mu^2(\tau)} \Bigg].
\end{aligned} \tag{2.13}$$

In particular,

$$\begin{aligned}
\mathbb{E}[1\!\!1_{\{\sigma > t\}} | \mathcal{D}_t^\tau] = 1\!\!1_{\{\tau > t\}} + 1\!\!1_{\{\tau \leq t\}} \Bigg[&\exp\big(-\mu^1(\tau)(t-\tau)\big) \frac{\mu^2(0) - \int_0^t \varphi(\tau\wedge s)ds}{\mu^1(\tau)+\mu^2(\tau)} \\
&+ \exp\big(-\mu^2(\tau)(t-\tau)\big) \frac{\mu^1(0) - \int_0^t \varphi(\tau\wedge s)ds}{\mu^1(\tau)+\mu^2(\tau)} \Bigg].
\end{aligned} \tag{2.14}$$

Proof. By (1.31), we have

$$\mathbb{E}[1\!\!1_{\{\sigma > T\}} | \mathcal{D}_t^\tau] = 1\!\!1_{\{\tau > t\}} \frac{\mathbb{P}(\sigma > T, \tau > t)}{\mathbb{P}(\tau > t)} + 1\!\!1_{\{\tau \leq t\}} \frac{\partial_s \mathbb{P}(\sigma > T, \tau > s)}{\partial_s \mathbb{P}(\tau > s)}\Big|_{s=\tau}. \tag{2.15}$$

Since $\mathbb{P}(\sigma > T, \tau > s) = \mathbb{P}(\tau_1 > T, \tau_2 > s) + \mathbb{P}(\tau_1 > s, \tau_2 > T) - \mathbb{P}(\tau_1 > T, \tau_2 > T)$, by applying (2.6) and (2.9), we obtain the proposition. □

Remark 2.1.10 By (2.13), we notice that conditioned on the first default, the second default time follows no longer an exponential distribution in general. The conditional probability is a linear combination of two exponential functions multiplied by some associated functions. Under very special condition that τ_1 and τ_2 follows independent identical law, i.e. $\mu^1(0) = \mu^2(0)$ and $\varphi = 0$, the second default time still follows the exponential law. Hence, we see that it is inappropriate to suppose the exponential stationary property to study successive defaults.

2.1.4 Explicit examples and numerical results

In this subsection, we study two explicit examples of the above illustrative model. We present some numerical results to show explicitly the correlation between the two defaults and the contagious jumps after the first default. Recall that the function φ characterizes the correlation between two credits.

For each example, we present two figures. The first figure shows the linear correlation between $1\!\!1_{\{\tau_1 > T^*\}}$ and $1\!\!1_{\{\tau_2 > T^*\}}$, i.e.

$$\rho = \frac{\mathrm{cov}\left[1\!\!1_{\{\tau_1 > T^*\}}, 1\!\!1_{\{\tau_2 > T^*\}}\right]}{\sqrt{\mathrm{Var}\left[1\!\!1_{\{\tau_1 > T^*\}}\right]}\sqrt{\mathrm{Var}\left[1\!\!1_{\{\tau_2 > T^*\}}\right]}}.$$

The second figure shows the contagious jump phenomenon. The reported quantity is the "jump" size of the so-called implied hazard rate $b^i(t,T)$ of the surviving credit when there is no interest rate and no recovery rate. To be more precise,

$$\begin{aligned} b^i(t,T) &= -\frac{1}{T-t}\ln \mathbb{E}[1\!\!1_{\{\tau_i > T\}}|\mathcal{D}_t] \\ &= \mu^i(0) - \int_0^{t\wedge\tau}\varphi(s)ds - 1\!\!1_{\{\tau\le t\}}\frac{1}{T-t}\ln\left(1 - \frac{\varphi(\tau)(T-t)}{\mu^j(0) - \varphi(\tau)(t-\tau) - \int_0^\tau \varphi(s)ds}\right). \end{aligned} \tag{2.16}$$

Before the first default, it equals $\mu^i(t)$. After the first default occurs, by (2.7), we observe that if $\varphi > 0$, there exists a positive jump at the first default time $t = \tau$ since $0 < \frac{\varphi(\tau)(T-\tau)}{\mu^j(0) - \int_0^\tau \varphi(s)ds} \le 1$, which implies that

$$\Delta b^i(\tau,T) = -\frac{1}{T-\tau}\ln\left(1 - \frac{\varphi(\tau)(T-\tau)}{\mu^j(0) - \int_0^\tau \varphi(s)ds}\right) > 0.$$

In the following, we fix $\mu^1(0) = \mu^2(0) = 0.01$, $T^* = 5$ years.

Example 2.1.11 $\varphi(t) = \alpha$ where α is a constant. By (2.6),

$$\mathbb{P}(\tau_1 > t_1, \tau_2 > t_2) = \exp\left(-\mu^1(0)t_1 - \mu^2(0)t_2 + \alpha t_1 t_2\right).$$

where α satisfies by (2.7) and (2.8)

$$-\mu^1(0)\mu^2(0) \leq \alpha \leq \frac{\min(\mu^1(0), \mu^2(0))}{T^*}.$$

The implied hazard rate is given by

$$b^i(t,T) = \mu^i(0) - \alpha(t \wedge \tau) - 1\!\!1_{\{\tau_j \leq t\}} \frac{1}{T-t} \ln\left(1 - \frac{\alpha(T-t)}{\mu^j(0) - \alpha t}\right)$$

and the "jump" at the default is given by

$$\Delta b^i(\tau, T) = -\frac{1}{T-\tau} \ln\left(1 - \frac{\alpha(T-\tau)}{\mu^j(0) - \alpha\tau}\right).$$

Figure 2.1 illustrates Example 2.1.11. α satisfies $-0.01\% \leq \alpha \leq 0.2\%$. In this example, the jump size increases with the correlation, also with the first default time. Of the two results, the former is quite natural. We explain the latter by a compensation effect since $\mu^i(t)$ decreases with time when there is no default event. We notice in addition that the correlation ρ is a linear function w.r.t. α.

Example 2.1.12 $\varphi(t) = \alpha \exp(-\alpha t) \min(\mu^1(0), \mu^2(0))$, where α is a constant parameter. Then, when $\alpha \neq 0$,

$$\begin{aligned}&\ln \mathbb{P}(\tau_1 > t_1, \tau_2 > t_2) = -(\mu^1(0)t_1 + \mu^2(0)t_2) \\ &+ (\mu^1(0) \wedge \mu^2(0)) \left(-e^{-\alpha(t_1 \wedge t_2)}|t_1 - t_2| + (t_1 + t_2) + \frac{2}{\alpha}(e^{-\alpha(t_1 \wedge t_2)} - 1)\right)\end{aligned}$$

When $\alpha \to 0$, we take the limit and get

$$\lim_{\alpha \to 0} \ln \mathbb{P}(\tau_1 > t_1, \tau_2 > t_2) = -\mu^1(0)t_1 - \mu^2(0)t_2,$$

which corresponds to the independence case. By (2.16), we have

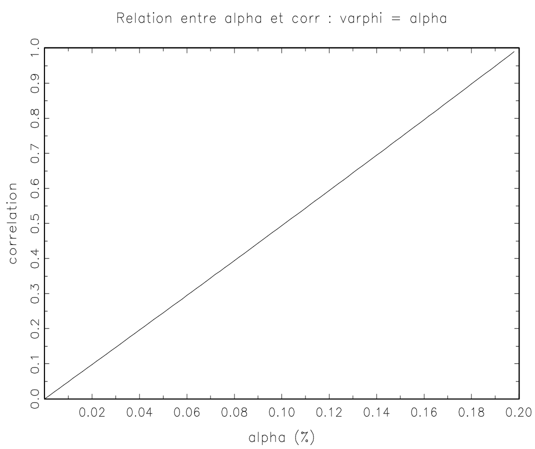
$$\begin{aligned}b^i(t,T) &= \mu^i(0) - (\mu^1(0) \wedge \mu^2(0))(1 - e^{-\alpha t}) \\ &- 1\!\!1_{\{\tau_j \leq t\}} \frac{1}{T-t} \ln\left(1 - \frac{(T-t)\alpha e^{-\alpha t}(\mu^1(0) \wedge \mu^2(0))}{\mu^j(0) - (\mu^1(0) \wedge \mu^2(0))[e^{-\alpha\tau}(\alpha t - \alpha\tau - 1) + 1]}\right)\end{aligned}$$

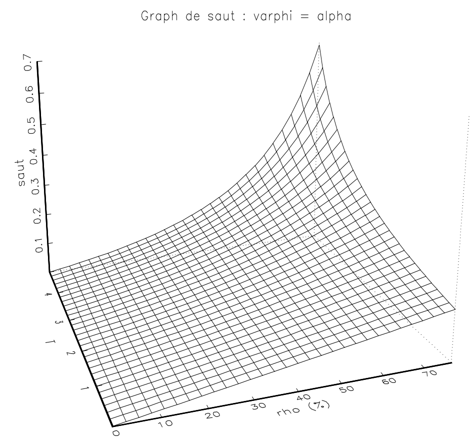
and

$$\Delta b^i(\tau, T) = -\frac{1}{T-\tau} \ln\left(1 - \frac{(T-\tau)\alpha e^{-\alpha\tau}(\mu^1(0) \wedge \mu^2(0))}{\mu^j(0) - (\mu^1(0) \wedge \mu^2(0))[1 - e^{-\alpha\tau}]}\right)$$

Notice that when $\mu^1(0) = \mu^2(0)$, the jump size $\Delta b^i(\tau, T) = -\frac{1}{T-\tau} \ln\left(1 - \alpha(T-\tau)\right)$.

Figure 2.1: The contagious jump and the linear correlation in Example 2.1.11 with $\mu^1(0) = \mu^2(0) = 0.01$, $T^* = 5$ years.





We now search for the bounds for α. Suppose first that $\alpha \geq 0$, in this case, we need only check the criterion (2.7). Since $\int_0^t \varphi(s)ds \leq \int_0^T \varphi(s \wedge t)ds$, it suffices that α satisfies

$$\int_0^T \varphi(s \wedge t)ds \leq \min(\mu^1(0), \mu^2(0)), \quad \forall 0 \leq t \leq T^*,$$

which implies $\int_0^{T^*} \varphi(s \wedge 0)ds \leq \min(\mu^1(0), \mu^2(0))$ since $\varphi(t)$ is decreasing. We then get

$$\alpha \leq \frac{1}{T^*}.$$

When α is negative, (2.7) is always true. We shall check (2.8). Notice that for any $\alpha < 0$, the left side of (2.8), i.e. $\alpha e^{-\alpha(t_1 \wedge t_2)} \min(\mu^1(0), \mu^2(0))$ is fixed when given $t_1 \wedge t_2$. However, the larger the value of $t_1 \vee t_2$, the smaller becomes the right side. So we need only consider the case where $t_1 = t_2$. We may suppose that $\mu^1(0) = \min(\mu^1(0), \mu^2(0))$, then,

$$\mu^1(0)\alpha e^{-\alpha t} \geq -\mu^1(0)\left(1 - \alpha \int_0^t e^{-\alpha s} ds\right)\left(\mu^2(0) - \alpha\mu^1(0)\int_0^t e^{-\alpha s} ds\right)$$

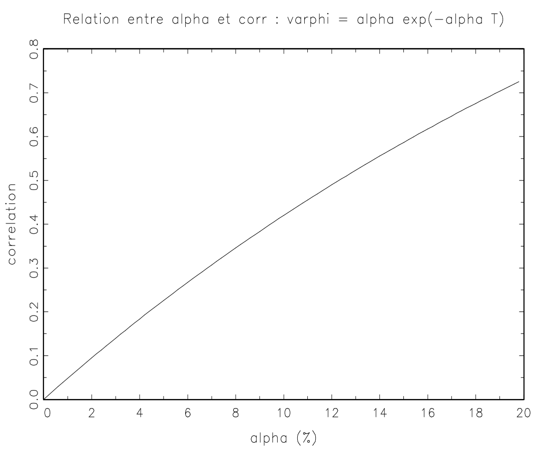
which follows

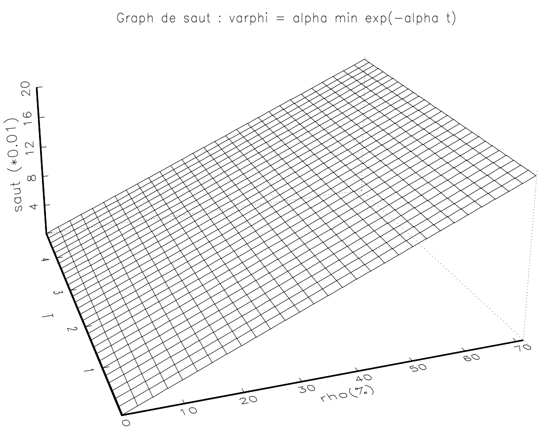
$$\alpha = -(\mu^2(0) + \mu^1(0)(e^{-\alpha t} - 1)) \geq -\mu^2(0).$$

By symmetry, we also have $\alpha \geq -\mu^1(0)$. Combining the two cases, we obtain α satisfies

$$-\max(\mu^1(0), \mu^2(0)) \leq \alpha \leq \frac{1}{T^*}.$$

Figure 2.2: illustrates example (2.1.12), α satisfies $-1\% \leq \alpha \leq 20\%$. We note that ρ reaches an upper limit of about 74% in this case.



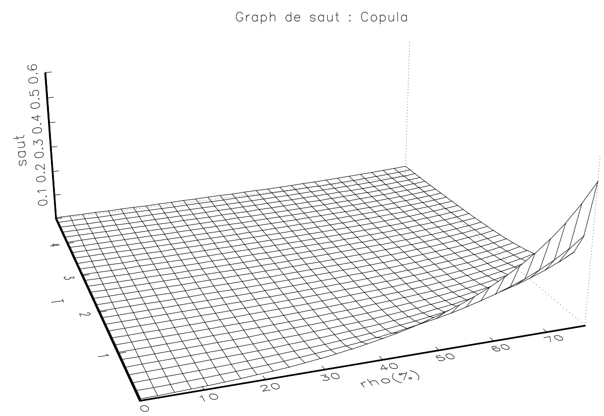


To compare the contagious jumps obtained by the Gaussian copula function. We present Figure 2.3 which shows the jump size of the defaultable bond price of the surviving credit at the first default time calculated by the Gaussian copula model. We notice that for a given correlation level, the jump size is decreasing with respect to the first default time. The practitioners find no obvious reason to support this phenomenon. In fact, Rogge and Schönbucher [70] has pointed out this disadvantage of the Gaussian copula. The authors propose to use Archimedean type copula functions to avoid this undesired property.

Table 2.1.13 Comparison of the joint survival probabilities by Example 2.1.11, Example 2.1.12 and the Gaussian copula. In each of the following table, the reported quantities are the joint probability

$$p_{ij} = \mathbb{P}(\tau_1 \in]t_i, t_{i+1}], \tau_2 \in]t_j, t_{j+1}]), (i, j = 0, 1, ..., T^*).$$

Figure 2.3: Contagious jumps after the first default by the Gaussian copula model. $\mu^1(0) = \mu^2(0) = 0.01$, $T^* = 5$.



In the three cases, we suppose that the individual default times follow the exponential distribution with parameters $\mu_1^0 = 10\%$ and $\mu_2^0 = 1\%$, which means that the sum of one column or one line in the three matrix are the same. The correlation parameters are chosen as follows. Let $\alpha = 0.1\%$ in Example 2.1.11, we then obtain the first matrix with $\mathbb{P}(\tau_1 > 5, \tau_2 > 5) = 0.59$, which implies the linear correlation $\rho = \rho(1\!\!1_{\{\tau_1 > T^*\}}, 1\!\!1_{\{\tau_2 > T^*\}}) = 13.88\%$ in the Gaussian copula model. Then we choose parameter in Example 2.1.12 to match this value of ρ.

Results by Example 2.1.11, $\alpha = 0.1\%$:

0.184%	0.182%	0.179%	0.177%	0.174%	8.621%
0.158%	**0.156%**	**0.154%**	**0.152%**	**0.150%**	**7.839%**
0.136%	**0.134%**	**0.132%**	**0.131%**	**0.129%**	**7.129%**
0.116%	**0.115%**	**0.113%**	**0.112%**	**0.111%**	**6.483%**
0.098%	**0.098%**	**0.097%**	**0.096%**	**0.095%**	**5.895%**
0.303%	**0.301%**	**0.299%**	**0.298%**	**0.297%**	**59.156%**

Results by Example 2.1.12, $\alpha = 12.12\%$:

0,199%	0,194%	0,191%	0,189%	0,186%	8,558%
0,168%	**0,158%**	**0,154%**	**0,152%**	**0,150%**	**7,829%**
0,144%	**0,133%**	**0,125%**	**0,122%**	**0,121%**	**7,146%**
0,122%	**0,114%**	**0,106%**	**0,099%**	**0,097%**	**6,511%**
0,103%	**0,097%**	**0,091%**	**0,084%**	**0,079%**	**5,924%**
0,258%	**0,289%**	**0,308%**	**0,319%**	**0,323%**	**59,156%**

Results by Gaussian copula, $\rho = 13,88\%$:

0.352%	0.282%	0.253%	0.233%	0.218%	8.177%
0.165%	**0.157%**	**0.151%**	**0.145%**	**0.141%**	**7.851%**
0.110%	**0.111%**	**0.110%**	**0.107%**	**0.106%**	**7.247%**
0.078%	**0.083%**	**0.084%**	**0.084%**	**0.083%**	**6.636%**
0.059%	**0.065%**	**0.066%**	**0.067%**	**0.067%**	**6.056%**
0.230%	**0.287%**	**0.312%**	**0.328%**	**0.341%**	**59.156%**

Remark 2.1.14 1. From the above tables, we see the different correlation structure in the three cases. In some way, our method can be viewed as a copula method because a particular form of joint probability matrix is specified.

2. From the bold part of each matrix, we can derive the conditional probabilities, for example, for any $i, j \geq 1$,

$$\mathbb{P}\big(\tau_1 \in (t_i, t_{i+1}], \tau_2 \in (t_j, t_{j+1}] \mid \tau_1 \geq 1, \tau_2 \geq 1\big) = \frac{\mathbb{P}\big(\tau_1 \in (t_i, t_{i+1}], \tau_2 \in (t_j, t_{j+1}]\big)}{\mathbb{P}(\tau_1 \geq 1, \tau_2 \geq 1)}.$$

and then $\mathbb{P}(\tau_k \leq T | \tau_1 \geq 1, \tau_2 \geq 1), (k = 1, 2)$ for integers $2 \leq T \leq T^*$. With a simple calculation, we see that the conditional marginal distribution remains in the exponential family with our model. However, this is not the case with the Gaussian copula model, which means that the correlation structure obtained today is not coherent with the conditional probability of tomorrow. Therefore, it's impossible to achieve a robust hedging strategy with the Gaussian copula model.

2.2 Default times in the general framework: the case of two credits

2.2.1 Compensator processes

Notations

We now consider the two-credits case in the general framework. Recall that the global information is represented by the filtration $\mathbb{G}$. Let τ_1 and τ_2 be two $\mathbb{G}$-stopping times

representing two default times. We denote by $\mathbb{D}^1$ (resp. $\mathbb{D}^2$) the filtration generated by the default process of τ_1 (resp. τ_2) as previously defined and by $\mathbb{D} = \mathbb{D}^1 \vee \mathbb{D}^2$. Let $\mathbb{F}$ be a subfiltration of $\mathbb{G}$ such that $\mathbb{G} = \mathbb{F} \vee \mathbb{D}$.
For convenience of writing, we introduce $\mathbb{G}^i = \mathbb{F} \vee \mathbb{D}^i$. We also consider the ordered default times. Let

$$\tau = \min(\tau_1, \tau_2) \quad \text{and} \quad \sigma = \max(\tau_1, \tau_2).$$

We define the associated filtrations $\mathbb{D}^\tau$ and $\mathbb{D}^\sigma$ respectively. Denote by $\mathbb{G}^\tau = \mathbb{F} \vee \mathbb{D}^\tau$ and by $\mathbb{G}^{\tau,\sigma} = \mathbb{F} \vee \mathbb{D}^\tau \vee \mathbb{D}^\sigma$.

The first default

It's convenient to work with the **Minimal Assumption** in the multi-credits case. It is clear that $(\mathbb{F}, \mathbb{G}^\tau, \tau)$ satisfies MA. Therefore, almost all the results deduced in the previous chapter can be applied directly to the first default time τ.

We know that the $\mathbb{G}$-compensator process Λ^i $(i = 1, 2)$ of τ_i exists and is unique. We now consider the $\mathbb{G}$-compensator Λ^τ of the first default time τ, which is stopped at τ. The following result has been given in Duffie [27] and been discussed in Jeanblanc and Rutkowski [54].

Proposition 2.2.1 *Let Λ^1, Λ^2 and Λ^τ be the $\mathbb{G}$-compensator processes of τ_1, τ_2 and τ respectively. Suppose that $\mathbb{P}(\tau_1 = \tau_2) = 0$, then*

$$\Lambda_t^\tau = \Lambda_{t\wedge\tau}^1 + \Lambda_{t\wedge\tau}^2.$$

Moreover, there exists $\mathbb{F}$-predictable processes $\Lambda^{i,\mathbb{F}}$ such that $\Lambda_{\tau\wedge t}^i = \Lambda_{\tau\wedge t}^{i,\mathbb{F}}$. Let $\Lambda^{\tau,\mathbb{F}} = \Lambda^{1,\mathbb{F}} + \Lambda^{2,\mathbb{F}}$, then $(\Lambda_{\tau\wedge t}^{\tau,\mathbb{F}}, t \geq 0)$ coincide with the compensator process of τ.

Proof. We first notice the equality $1\!\!1_{\{\tau \leq t\}} = 1\!\!1_{\{\tau_1 \leq \tau\wedge t\}} + 1\!\!1_{\{\tau_2 \leq \tau\wedge t\}} - 1\!\!1_{\{\tau_1 = \tau_2 \leq t\}}$. Since $1\!\!1_{\{\tau_1 = \tau_2\}} = 0$ a.s., we know that $1\!\!1_{\{\tau \leq t\}} = 1\!\!1_{\{\tau_1 \leq \tau\wedge t\}} + 1\!\!1_{\{\tau_2 \leq \tau\wedge t\}}$ a.s.. On the other hand, $(1\!\!1_{\{\tau_1 \leq \tau\wedge t\}} - \Lambda_{\tau\wedge t}^1, t \geq 0)$ and $(1\!\!1_{\{\tau_2 \leq \tau\wedge t\}} - \Lambda_{\tau\wedge t}^2, t \geq 0)$ are $\mathbb{G}$-martingales, by taking the sum, we have $\big(1\!\!1_{\{\tau \leq t\}} - (\Lambda_{\tau\wedge t}^1 + \Lambda_{\tau\wedge t}^2), t \geq 0\big)$ is also a $\mathbb{G}$-martingale. Finally, it suffices to note that $(\mathbb{F}, \mathbb{G}, \tau)$ satisfies MA to end the proof. □

Remark 2.2.2 With the same method, we can recover the result of [27] which confirms that for $\mathbb{G}$-stopping times $\tau_1, \cdots, \tau_n$ and $\tau = \min(\tau_1, \cdots, \tau_n)$ whose $\mathbb{G}$-compensator are $\Lambda^1, \cdots, \Lambda^n$ and Λ^τ respectively, if $\mathbb{P}(\tau_i = \tau_j) = 0$ for any $1 \leq i < j \leq n$, then $\Lambda_t^\tau = \sum_{i=1}^n \Lambda_{\tau\wedge t}^i$. In addition, we can relax the condition to be $\mathbb{P}(\tau = \tau_i = \tau_j) = 0$.

The second default

To study the compensator process of the second default time σ, we shall use some results already established in the previous chapter. In fact, since $\mathbb{G}^{\tau,\sigma} = \mathbb{G}^{\tau} \vee \mathbb{D}^{\sigma}$, we know that MA also holds for $(\mathbb{G}^{\tau}, \mathbb{G}^{\tau,\sigma}, \sigma)$. This observation enables us to deduce some properties of the second default time with the filtration generated by the first default time $\mathbb{G}^{\tau}$. In particular, we know that there exists some $\mathbb{G}^{\tau}$-predictable process $\Lambda^{\sigma,\mathbb{G}^{\tau}}$ which coincides with the $\mathbb{G}^{\tau,\sigma}$-compensator process Λ^{σ} of the second default time σ, i.e. $\Lambda^{\sigma}_t = \Lambda^{\sigma,\mathbb{G}^{\tau}}_{\sigma\wedge t}$. The calculation of Λ^{σ} is easy on $\{\tau > t\}$. In fact, we have the following result.

Proposition 2.2.3 *Let $\sigma = \max(\tau_1, \tau_2)$ and let Λ^{σ} be the $\mathbb{G}^{\tau,\sigma}$-compensator process of σ. If $\mathbb{P}(\tau_1 = \tau_2) = 0$, then $\Lambda^{\sigma}_{\tau\wedge t} = 0$.*

Proof. By definition, $(1\!\!1_{\{\sigma\leq t\}} - \Lambda^{\sigma}_t, t \geq 0)$ is a $\mathbb{G}$-martingale. Then $(1\!\!1_{\{\sigma\leq \tau\wedge t\}} - \Lambda^{\sigma}_{\tau\wedge t}, t \geq 0)$ is also a $\mathbb{G}$-martingale. Since $\mathbb{P}(\sigma = \tau) = 0$, we know that $1\!\!1_{\{\sigma\leq \tau\wedge t\}} = 0$, a.s., which implies that $\Lambda^{\sigma}_{\tau\wedge t} = 0$. $\square$

The calculation of Λ^{σ} on $\{\tau \leq t\}$ is more complicated. We first recall that in the single-credit case, the process G and its Doob-Meyer decomposition plays an important role. By analogy, we now introduce the process H defined by $H_t = \mathbb{P}(\sigma > t|\mathcal{G}^{\tau}_t)$ and we shall discuss its property. By condition MA and applying directly Theorem 1.3.4, we have the following result and we know that H is important to calculate the compensator process of σ.

Proposition 2.2.4 *Let $H_t = \mathbb{P}(\sigma > t|\mathcal{G}^{\tau}_t)$ and $H = M^H - A^H$ be the Doob-Meyer decomposition of H where M^H is a $\mathbb{G}^{\tau}$-martingale and A^H is an increasing $\mathbb{G}^{\tau}$-predictable process. Then we have*

$$dA^H_t = H_{t-} d\Lambda^{\sigma,\mathbb{G}^{\tau}}_t.$$

By discussions on the case before and especially after the stopping time τ in the previous chapter, H can be calculated explicitly. The only hypothesis we need is that the conditional joint probability has a density, as introduced below.

Hypothesis 2.2.5 We suppose that the conditional joint probability $\mathbb{P}(\tau > u, \sigma > v|\mathcal{F}_t)$ of the ordered default times τ and σ admits a density p_t, that is

$$\mathbb{P}(\tau > u, \sigma > v|\mathcal{F}_t) = \int_u^{\infty} d\theta_1 \int_v^{\infty} d\theta_2 p_t(\theta_1, \theta_2). \tag{2.17}$$

Remark 2.2.6 Note that $\int_{\theta}^{\infty} p_t(\theta, v)dv = \alpha^{\tau}_t(\theta)$ where $\alpha^{\tau}_t(\theta)$ is the density of $G^{\theta}_t = \mathbb{P}(\tau > \theta|\mathcal{F}_t)$.

For the above joint probability, we only need to consider the case where $u < v$ since $\tau \le \sigma$. Otherwise, it suffices to consider the marginal conditional probability $G_t^u = \mathbb{P}(\tau > u|\mathcal{F}_t)$ characterized by its density $\alpha_t^\tau(\theta)$, i.e. $\mathbb{P}(\tau > u|\mathcal{F}_t) = \int_u^\infty \alpha_t^\tau(\theta)d\theta$. Moreover, for any $\theta_1 > \theta_2$, we have $p_t(\theta_1, \theta_2) = 0$. We shall see in the following that using the martingale density $p_t(\theta_1, \theta_2)$, many calculations are similar as in the previous chapter.

Proposition 2.2.7 *We assume that Hypothesis 2.2.5 holds. Then the $\mathbb{G}^\tau$-supermartingale H is calculated by*

$$H_t = 1\!\!1_{\{\tau>t\}} + 1\!\!1_{\{\tau\le t\}} \frac{\int_t^\infty p_t(\tau, v)dv}{\alpha_t^\tau(\tau)} = 1 - \frac{\int_{\tau\wedge t}^t p_t(\tau\wedge t, v)dv}{\alpha_t^\tau(\tau\wedge t)}. \tag{2.18}$$

In addition, H is continuous.

Proof. It is obvious that $H_t 1\!\!1_{\{\tau>t\}} = 1\!\!1_{\{\tau>t\}}$. Using (1.31) and the density $p_t(\theta_1, \theta_2)$, we have

$$\mathbb{E}[1\!\!1_{\{\sigma>t\}}|\mathcal{G}_t^\tau]1\!\!1_{\{\tau\le t\}} = \frac{\partial_s \mathbb{P}(\sigma > t, \tau > s|\mathcal{F}_t)}{\partial_s \mathbb{P}(\tau > s|\mathcal{F}_t)}\Big|_{s=\tau} 1\!\!1_{\{\tau\le t\}} = \frac{\int_t^\infty p_t(\tau, v)dv}{\int_\tau^\infty p_t(\tau, v)dv},$$

which implies (2.18). Moreover, $H_\tau = 1$, which means that H is continuous. $\square$

We also consider the family of the $\mathbb{G}^\tau$-martingales

$$\left(H_t^\theta = \mathbb{P}(\sigma > \theta|\mathcal{G}_t^\tau), t \ge 0\right) \quad \text{for any} \quad \theta \ge 0.$$

Both cases where $\theta \ge t$ or $\theta < t$ are important here. The calculation of H^θ is also similar by using (1.31) and the result is given below.

Proposition 2.2.8 *Under Hypothesis 2.2.5, we have*

$$\begin{aligned} H_t^\theta &= 1\!\!1_{\{\tau>t\}} \frac{\mathbb{P}(\tau > t, \sigma > \theta|\mathcal{F}_t)}{\mathbb{P}(\tau > t|\mathcal{F}_t)} + 1\!\!1_{\{\tau\le t\}} \frac{\partial_s \mathbb{P}(\sigma > \theta, \tau > s|\mathcal{F}_t)}{\partial_s \mathbb{P}(\tau > s|\mathcal{F}_t)}\Big|_{s=\tau} \\ &= 1\!\!1_{\{\tau>t\}} \frac{\int_t^\infty du \int_{\theta\vee u}^\infty p_t(u, v)dv}{\int_t^\infty du \int_u^\infty p_t(u, v)dv} + 1\!\!1_{\{\tau\le t\}} \frac{\int_{\tau\vee\theta}^\infty p_t(\tau, v)dv}{\int_\tau^\infty p_t(\tau, v)dv} \end{aligned} \tag{2.19}$$

In particular, we have for any $T > t$,

$$H_t^T = 1\!\!1_{\{\tau>t\}} \frac{\int_t^\infty du \int_{T\vee u}^\infty p_t(u, v)dv}{\int_t^\infty du \int_u^\infty p_t(u, v)dv} + 1\!\!1_{\{\tau\le t\}} \frac{\int_T^\infty p_t(\tau, v)dv}{\int_\tau^\infty p_t(\tau, v)dv}$$

As an immediate consequence of MA and the above proposition, we have the following result.

Corollary 2.2.9 *For any $T > t$, the process defined by*

$$\left(1\!\!1_{\{\tau>t\}} \frac{\int_t^\infty du \int_{T\vee u}^\infty p_t(u,v)dv}{\int_t^\infty du \int_u^\infty p_t(u,v)dv} + 1\!\!1_{\{\tau\le t,\sigma>t\}} \frac{\int_T^\infty p_t(\tau,v)dv}{\int_t^\infty p_t(\tau,v)dv}, \quad t \ge 0 \right)$$

is a $\mathbb{G}^{\tau,\sigma}$-martingale.

Proof. It suffices to note that under MA for $(\mathbb{G}^\tau, \mathbb{G}^{\tau,\sigma}, \sigma)$, the conditional survival probability is given by $\mathbb{P}(\sigma > T|\mathcal{G}_t^{\tau,\sigma}) = 1\!\!1_{\{\sigma>t\}} \frac{H_t^T}{H_t}$, which follows immediately the corollary by (2.18) and (2.19). $\square$

Remark 2.2.10 If the filtration $\mathbb{F}$ is trivial, or in other words, if $\mathbb{G} = \mathbb{D}$, then the density function $p(\theta_1, \theta_2)$ does not depend on t. Hence

$$H_t = 1\!\!1_{\{\tau>t\}} + 1\!\!1_{\{\tau\le t\}} \frac{\int_t^\infty p(\tau,v)dv}{\int_\tau^\infty p(\tau,v)dv}$$

is absolutely continuous. Since H is a $\mathbb{D}^\tau$-supermartingale, it is decreasing. Proposition 2.2.4 implies that the $\mathbb{D}^\tau$-compensator process $\Lambda^{\sigma,\mathbb{D}^\tau}$ is given by $\Lambda_t^{\sigma,\mathbb{D}^\tau} = -\ln H_t$. The general case is discussed below.

The following theorem is the main result of this section which gives the compensator process of σ. We see that Theorem 1.5.7 can be applied without much difficulty to the multi-credits case. One important point to note is that $(\mathbb{G}^\tau, \mathbb{G}^{\tau,\sigma}, \sigma)$ satisfy MA.

Theorem 2.2.11 *We assume that Hypothesis 2.2.5 holds. Then*

1) the process $(H_t^\theta = \mathbb{P}(\sigma > \theta|\mathcal{G}_t^\tau), t \ge 0)$ admits a density $(\alpha_t^\sigma(\theta), t \ge 0)$, i.e. $H_t^\theta = \int_\theta^\infty \alpha_t^\sigma(s)ds$, which is given by

$$\alpha_t^\sigma(\theta) = 1\!\!1_{\{\tau>t\}} \frac{\int_t^\infty du p_t(u,\theta)}{\int_t^\infty du \int_u^\infty p_t(u,v)dv} + 1\!\!1_{\{\tau\le t\}} \frac{p_t(\tau,\theta)}{\alpha_t^\tau(\tau)}; \tag{2.20}$$

2) the $\mathbb{G}^{\tau,\sigma}$-compensator process Λ^σ of σ is given by

$$d\Lambda_t^\sigma = 1\!\!1_{[\tau,\sigma]}(t) \frac{p_t(\tau,t)}{\int_t^\infty p_t(\tau,v)dv} dt. \tag{2.21}$$

Proof. 1) We obtain directly $\alpha_t^\sigma(\theta)$ by taking derivative of H_t^θ given by (2.19) with respect to θ.
2) Similar as in Theorem 1.5.7, we notice that the process

$$\left(H_t + \int_0^t \alpha_v^\sigma(v)dv = H_t + \int_0^t 1\!\!1_{[\tau,\infty[}(v) \frac{p_v(\tau,v)}{\alpha_v^\tau(\tau)} dv, t \ge 0 \right) \tag{2.22}$$

is a $\mathbb{G}^\tau$-martingale, which implies that $dA_t^H = \mathbb{1}_{\{\tau \le t\}} \frac{p_t(\tau,t)}{\alpha_t^\tau(\tau)}$. Since H is continuous, we obtain 2) directly by Proposition 2.2.4. □

Remark 2.2.12 1. It's not difficult to see that the framework we propose above can be extended directly to study the successive defaults. A natural application will be on the dynamic portfolio losses modelling in the pricing of CDO tranches since under standard market assumptions, the loss on a portfolio is determined by the number of defaulted credits. The key term is the conditional joint probability of the ordered default times with resepct to the filtration $\mathbb{F}$.

2. Indeed, we here propose an original point of view for the credit correlation anaylysis: to study the ordered default times rather than the individual default times and their joint distribution by the copula models. The difficulty of the latter approach lies in the incompatibility between the dynamic property of the marginal distributions and the static property of the copula functions, which has been shown in the illustrative example. On the contrary, for the ordered-defaults, we deduce in the general way and we think it's a promising framework which deserves further studies.

2.3 Appendix

2.3.1 Copula models

In this section, we review the copula model which are widely used in the credit correlation modelling by the market practitioners. Li [61] first introduced the copula method to the credit dependency analysis. Schönbucher and Schubert [74] studied the dynamics properties implied by the copula model. The copula model are often combined with the reduced form approach to characterize the default correlation. The method is static, that is, the model is applied at each time for a given maturity. Then the procedure is repeated the next day.

We recall the definition of the copula function. For a detailed introduction to this method, we send the readers to the monograph of Nelsen [65].

Definition 2.3.1 A *n-dimensional copula* $\mathcal{C}$ is a function defined on $[0,1]^n$ and valued in $[0,1]$ which satisfies the following properties:

1) for any $(u_{1,1},\cdots,u_{n,1})$ and $(u_{1,2},\cdots,u_{n,2})$ with $u_{k,1} \le u_{k,2}$, $k=1,\cdots,n$, we have

$$\sum_{i_1=1}^{2}\cdots\sum_{i_n=1}^{2}(-1)^{i_1+\cdots+i_n}\mathcal{C}(u_{1,i_1},\cdots,u_{n,i_n}) \ge 0;$$

2) $\mathcal{C}(1, \cdots, 1, u, 1, \cdots, 1) = u$.

We see that a copula function defines a multi-dimensional probability distribution function whose marginal distributions are uniform on $[0, 1]$. Moreover, for a family of one-dimensional distribution functions $(F_1(x_1), \cdots, F_1(x_n)) \in [0, 1]^n$, $(x_1, \cdots, x_n) \rightarrow \mathcal{C}(F_1(x_1), \cdots, F_n(x_n))$ is a cumulative distribution function on $\mathbb{R}^n$. The relationship between the marginal distributions and their joint distributions is given by the Sklar's theorem.

Theorem 2.3.2 *(Sklar) Let F be an n-dimensional distribution function with continuous margins $F_1, ..., F_n$. Then F has a unique copula representation:*

$$F(x_1, ...x_n) = \mathcal{C}(F_1(x_1), ..., F_n(x_n)).$$

The popularity of the copula model lies in its efficiency to analyze separately the marginal distributions and the joint distribution and thus to deal with large size portfolios. In fact, from the Sklar's theorem, we deduce immediately a two-steps procedure to obtain the joint distribution function.

1) calculate the marginal survival probabilities $\mathbb{P}(\tau_i > T_i^*)$ with the quoted CDS spreads;

2) choose a copula function to obtain the joint survival distribution $\mathbb{P}(\tau_1 > T_1^*, \cdots, \tau_n > T_n^*)$ which satisfies some empirical conditions.

The Sklar's theorem implies the possibility to capture the "real" structure of the portfolio dependence by selecting a suitable copula. Moreover, the standard assumption is that the choice of the copula function is independent of the marginal distributions. Therefore, most discussions concentrate on the choice of a particular copula function. Frey and McNeil [33] studied the impact of different types of copula functions on the default correlation. Among the others, Rogge and Schönbucher [70] proposed the t-student or the Archimedean copula. We shall also note the one-factor Gaussian copula model by Andersen, Sidenius and Basu [1] which are used for the CDO pricing.

One important point of [74] is the analysis of the dynamic properties of the survival probabilities implied from a given copula function. In their model, the default is constructed as in the example presented previously where the (H)-hypothesis holds, $\tau_i = \inf\{t : \Phi_t^i \geq \xi_i\}$ $(i = 1, \cdots, n)$ where Φ^i is an F-adapted, continuous, increasing processes satisfying $\Phi_0^i = 0$ and $\Phi_\infty^i = +\infty$. and $\xi_1, \cdots, \xi_n$ are i.i.d exponential random variables with parameter 1 which are independent with the σ-algebra $\mathcal{F}_\infty$. Recall the survival probability $q_i(t) := \mathbb{P}(\tau_i > t|\mathcal{F}_\infty) = \mathbb{P}(\tau_i > t|\mathcal{F}_t) = e^{-\Phi_t^i}$. This model has also been discussed by Bielecki and Rutkowski [9] as an example of conditionally independent default times.

The correlation of defaults is imposed by introducing an n-dimensional copula function $\mathcal{C}(x_1, \cdots, x_n)$. The authors suppose that the choice of the copula function is independent of the marginal probability q_i and the joint survival probability is given by

$$G(t_1, \cdots, t_n) = \mathbb{P}(\tau_1 > t_1, \cdots, \tau_n > t_n \mid \mathcal{F}_\infty) = \mathcal{C}(q_1(t_1), \cdots, q_n(t_n)).$$

The main result of [74] shows that at any time $t > 0$, we can deduce, from the initial joint survival probability $G(t_1, \cdots, t_n)$, the conditional probabilities on all default scenarios, i.e. $\mathbb{P}(\tau_i > T | \mathcal{F}_\infty \vee \mathcal{D}_t)$ for any $T > t$. Hence, by choosing a static copula function, one obtains the dynamics of the conditional probabilities.

Remark 2.3.3 The idea of using the $\mathbb{F}$-copula function to calculate $\mathbb{G}$-conditional probabilities is similar with ours. However, our method is much more general. We don't need the (H)-hypothesis and by working with a martingale measure, we can calculate all $\mathbb{G}$-conditional expectations with explicite formulae.

Part II

Asymptotic Expansions by Stein's Method: A finance concerned result and a theoretical result

Chapter 3

Approximation of Large Portfolio Losses by Stein's Method and Zero Bias Transformation

The calculation of cumulative losses on large portfolios is a classical problem in the insurance and in the risk management to measure the counterpart risks in an aggregate way. Recently, it attracts new attentions of the market practitioners who search for pricing portfolio credit products such as CDOs. This problem is particularly difficult when the size of the portfolio is large, since in the high-dimensional case, the exact modelling of correlation structure is beyond analytical tractability. Therefore, both an acceptable modelling approach and rapid numerical simulations bring real challenges.

The market adopts a simplified approach, the factor models, to describe the correlation structure between the defaults, as proposed in Andersen, Sidenius and Basu [1] and Gregory and Laurent [43]. To be more precise, the cumulative loss L_T before a fixed date T is the sum of all individual losses, i.e. $L_T = \sum_{i=1}^{n} L_i Z_i(T)$ where L_i is the loss given default of each credit and $Z_i(T) = 1\!\!1_{\{\tau_i \leq T\}}$ indicates the occurrence of default of credit i before T. In the factor models, $Z_i(T)$ are supposed to be correlated through a common factor Y and conditioned on this factor, the defaults are supposed to be independent. From the theoretical point of view, we are hence interested in the sum of independent random variables, which is one of the most important subject in the probability theory. We know from the central limit theorem (CLT) that the total loss converges in law to the normal distribution when the size of the portfolio is large enough.

To the finance concern, Vasicek [80] first applies the normal approximation to an homogeneous portfolio of loans to achieve faster numerical computation. This method is extended to CDOs loss calculation by Shelton [75] for the non-homogeneous portfolios where the individual loss distribution is not necessarily identical, and eventually to CDO^2 portfolios where each component is also a CDOs tranche. The method of

[75] is based on the normal approximation by replacing the total loss of an exogenous portfolio by a normal random variable of the same expectation and the same variance. Some improvements have been proposed, including the large-deviations approximation by Dembo, Deuschel and Duffie [24] and by Glasserman, Kang and Shahabuddin [36]. Glasserman [35] compares several methods including the saddlepoint and Laplace approximations. In Antonov, Mechkov and Misirpashaev [2], the authors provide approximation corrections to the expectation of the CDOs payoff function using the saddle-point method, which consists of writing the expectation as the inverse Laplace transform and expanding the conditional cumulant generating function at some well-chosen saddle point. This method coincides with the normal approximation when taking expansion at one particular point and show in general better approximation results.

In the above papers, the authors give financial applications, but no discussion on the estimation of approximation errors is presented. The rate of convergence of CLT is given by the Berry-Esseen inequality. For example, for the binomial random variables, the rate of convergence is of order $\frac{1}{\sqrt{n}}$ for a fixed probability p. However, in the credit analysis. the approximation accuracy deserves thorough investigation since the default probabilities are usually small and the normal approximation fails to be robust when the size of portfolio n is fixed. This is the main concern of our work.

In this chapter, we provide, by combining the Stein's method and the zero bias transformation, a correction term for the normal approximation. The Stein's method, which is a powerful tool in proving the CLT and in studying the related problems, shall be presented in Subsections 3.1.3 and 3.3.1. The error estimation of corrected approximation is obtained when the solution of the associated Stein's equation has bounded third order derivative. In the binomial case, the error bound is of order $O(\frac{1}{n})$. It is shown that the corrector vanishes for symmetrically distributed random variables. For asymmetric cases such as the binomial distribution with very small probability, we obtain the same accuracy of approximation after correction. In addition, the summand variables are not required to be identically distributed. The result is then extended to the "call" function which is essential to the CDOs evaluation. Since this function is not second-ordered derivable, the error estimation is more complicated. The principal tool is a concentration inequality.

We then apply the result to calculate the conditional losses of the CDOs tranches and numerical tests preform well in comparison with other approximations available. The impact of correlation and the common factor is studied. Our method gives better approximations when the correlation parameter is small. The method is less efficient when the defaults are strongly correlated, which means, in the normal factor model framework, that the conditional loss calculation is almost normal distributed. In this case, there is effectively no need of correction.

This chapter is organized as follows. We first review briefly some convergence results for the sum of independent random variables and the literature dedicated to the Stein's method and the zero bias transformation. Section 3.2 and Section 3.3 are devoted to estimations results in these two contexts respectively. Section 3.2 gives estimation on the difference between the sum random variable and its zero bias transformation. Section 3.3 deals with the growing speed of the auxiliary function which is the solution of the Stein's equation. The main results are given in Section 3.4 where we present an approximation corrector and we estimate the approximation error. The call function has been studied in particular since it lacks regularity. Numerical examples are presented to show the approximation result. The last section concentrates on the application to the CDO loss approximation.

3.1 Introduction

In this introductory section, we first present the CDO and the factor model which is the basic model in the following this chapter. Second, We give a brief review of the sum of independent random variables and some well-known results concerning the convergence in the central limit theorem. This is of course a classical subject. However, we shall re-discuss it in our conditional loss context. Finally, we give an overview of literature of the Stein's method, which is, combined with the zero bias transformation, our main tool to treat the problem. Some useful properties shall be developed in the next section.

3.1.1 Factor models for CDOs

A CDO is a special transaction between investors and debt issuers through the intervention of a special purposed vehicle (SPV). The CDO structure serves as an efficient tool for banks to transfer and control their risks as well as decrease the regulatory capital. On the other hand, this product provides a flexible choice for investors who are interested in risky assets but have constrained information on each individual firm. A CDO contract consists of a reference portfolio of defaultable assets such as bonds (CBO), loans (CLO) or CDS (synthetic CDO). A CDO contract deals in general with large portfolios of 50 to 500 firms. Instead of the individual assets of the portfolio, one can invest in the specially designed notes based on the portfolio according to his own risk preference. These notes are called tranches. There exist in general the *equity tranche*, the *mezzanine tranches* and the *senior tranche*, and we suppose that the nominal amounts of tranches are denoted by N_E, N_M and N_S. The cash-flow of a tranche consists of interest and principal repayments which obey prioritization policy: the senior CDO tranche which carries the least interest rate is paid first, then follow the lower subordinated mezzanine tranches and at last the equity tranche which carries excess interest rate. More precisely, the repayment depends on the cumulative loss of the underlying portfolio which is given by $L_t = \sum_{i=1}^n N_i(1-R_i)1\!\!1_{\{\tau_i \le t\}}$ where N_i is the nominal value of each firm. If there is no default, all tranches are fully repayed. The first defaults only affect the equity tranche until the cumulative loss has arrived the total nominal amount of the equity tranche and the loss on the tranche is given by

$$L_t^E = L_t 1\!\!1_{[0,N_E]}(L_t) + N_E 1\!\!1_{]N_E,+\infty[}(L_t) \tag{3.1}$$

Notice that $L_t^E = L_t - (L_t - N_E)^+$, which is the difference of two "call" functions with strike values equal 0 and N_E. The following defaults will continue to hit the other tranches along their subordination orders. The loss on the mezzanine tranche and the senior tranche is calculated similarly as for the equity tranche. Hence, for the pricing of a CDO tranche, the call function plays an important role.

On the other hand, the market is experiencing several new trends recently, such as the creation of diversified credit indexes like Trac-X, iBoxx etc. A Trac-X index consists in general of 100 geographically grouped entreprises while each one is equally weighted in the reference pool. All these credits are among the most liquid ones on the market, so the index itself reflects flexibility and liquidity. The derivative products based on this index are rapidly developed. The CDOs of Trac-X are of the same characteristics of the classical ones with standard tranches as $[3\%, 6\%]$, $[6\%, 9\%]$, $[9\%, 12\%]$, $[12\%, 1]$, (or $[12\%, 22\%]$, $[22\%, 1]$) and the transaction of single tranche is possible.

We now present the *factor model*, which has become the standard model on the market for CDOs. The one-factor normal model has been first proposed by Andersen, Sidenius and Basu [1] and Gregory and Laurent [43]. It is a copula model which we introduce in Subsection 2.3.1.

We consider a static context and we neglect the filtration. The time horizon is fixed to be T. The default times $\tau_1, \cdots, \tau_n$ are defined as a special case in the Schönbucher and Schubert's model with the filtration $\mathbb{F}$ being trivial. To be more precise, the default time is defined as the first time that $q_i(t)$ reaches a uniformly distributed threshold U_i on $[0,1]$, i.e.

$$\tau_i = \inf\{t : q_i(t) \le U_i\},$$

where $q_i(t)$ is the expected survival probability up to time t. Clearly, $q_i(0) = 1$ and $q_i(t)$ is decreasing. Moreover, $q_i(t)$ can be calibrated from the market data for each credit.

The characteristic of the factor model lies in the correlation specification of the thresholds U_i. Let $Y_1, \cdots, Y_n$ and Y be independent random variables where Y represents a common factor characterizing the macro-economic impact on all firms and Y_i are idiosyncratic factors reflecting the financial situation of individual credits. Let

$$X_i = \sqrt{\rho_i}Y + \sqrt{1-\rho_i}Y_i$$

be the linear combination of Y and Y_i. The coefficient ρ_i is the weight on the common factor and thus the linear correlation between X_i and X_j is $\sqrt{\rho_i \rho_j}$. The default thresholds are defined by $U_i = 1 - F_i(X_i)$ where F_i is the cumulative distribution function of X_i. Then

$$\tau_i = \inf\{t : \sqrt{\rho_i}Y + \sqrt{1-\rho_i}Y_i \le F_i^{-1}(p_i(t))\}$$

where $p_i(t) = 1 - q_i(t)$ is the expected default probability of credit i before the time t. It is obvious that conditioned on the common factor Y, the defaults are independent.

The survival probability is

$$\mathbb{P}(\tau_i > t) = \mathbb{P}(\sqrt{\rho_i}Y + \sqrt{1-\rho_i}Y_i \ge F_i^{-1}(p_i(t))) = p_i(t).$$

Conditioned on the common factor Y, we have $\mathbb{P}(\tau_i \le t|Y) = F_i^Y\big(\frac{F_i^{-1}(p_i(t))-\sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\big)$ where F_i^Y is the distribution function of Y_i. In particular, in the normal factor case where Y and Y_i are standard normal random variables, X_i is also a standard normal random variable, then we have

$$p_i(t|Y) = \mathbb{P}(\tau_i \le t \mid Y) = \mathcal{N}\left(\frac{\mathcal{N}^{-1}(p_i(t)) - \sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\right). \tag{3.2}$$

We note that although the Gaussian factor model is very popular among the practitioners, it can be extended without much difficulty to models containing several factors which can follow any distribution.

Each default before the maturity T brings a loss to the portfolio. Then the total loss on the portfolio at maturity is given by

$$L_T = \sum_{i=1}^{n} N_i(1-R_i)\mathbb{1}_{\{\tau_i \le T\}}, \tag{3.3}$$

where N_i is the notional value of each credit i and R_i is the recovery rate. In the following, we suppose R_i is constant. Conditional on the common factor Y, we can rewrite

$$L_T = \sum_{i=1}^{n} N_i(1-R_i)\mathbb{1}_{\left\{Y_i \le \frac{F_i^{-1}(p_i(T))-\sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\right\}}.$$

Hence, the conditional total loss L_T on the factor Y can be written as the sum of independent Bernoulli random variables, each with probability $p_i(Y) = p_i(T|Y)$. In particular, for an homogenous portfolio where N_i, R_i and $p_i(Y)$ are equal, L_T is a binomial random variable.

The common factor Y follows certain distribution. Denote by $F(y) = \mathbb{P}(Y \le y)$ the distribution function of Y. Then for any function h, if we denote by $H(Y) = \mathbb{E}[h(L_T)|Y]$, we have

$$\mathbb{E}[h(L_T)] = \int_{\mathbb{R}} H(y)dF(y).$$

That is to say, we can study $\mathbb{E}[h(L_T)]$ in two successive steps. First, we consider the conditional expectation $\mathbb{E}[h(L_T)|Y]$ and second, we study the role played by the factor Y.

We recall that for a CDO tranche with the lower and upper barriers of the tranche A and B, the evaluation is determined by the loss on the tranche

$$L_T(A,B) = (L_T - A)^+ - (L_T - B)^+. \tag{3.4}$$

Notice that $L_T(A,B)$ is the call spread, i.e. the difference between two European call functions. Hence, we are interested in calculating the expectation of the call function $h(x) = (x-k)^+$.

3.1.2 Sum of independent random variables

As shown above, under the factor model framework, our first objective is to study the conditional losses. Since the defaults are conditionally independent, this step is equivalent to calculating the expectation of the call function for sum of independent Bernoulli random variables.

The sum of independent random variables is a very classical subject in the probability theory which is related to the law of large numbers and the central limit theorem. The most simple case is the sum of i.i.d. Bernoulli random variables which follows the Binomial distribution. Let S_n be a Binomial random variable with parameters (n,p) where $n \geq 1$ is a integer. Historically, Laplace proved that when $n \to +\infty$,

$$\mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx.$$

In other words, the sequence of random variables $\left(\frac{S_n - np}{\sqrt{np(1-p)}}\right)_{n\geq 1}$ converges in law to a standard normal distribution. This is the original form of central limit theorem.

The result of Laplace can be stated in the following way: Let $(\xi_i)_{i\geq 1}$ be a sequence of i.i.d. random variables of Bernouilli distribution with parameter p. Then S_n has the same law as $\xi_1 + \cdots + \xi_n$. For any integer $1 \leq i \leq n$, let $X_{n,i} = (\xi_i - p)/\sqrt{np(1-p)}$. Notice that $(X_{n,i})_{1\leq i\leq n}$ are i.i.d random variables. Denote by $W_n = X_{n,1} + \cdots + X_{n,n}$. Then the sequence $(W_n)_{n\geq 1}$ converges in law to the standard normal distribution. We observe immediately that the behavior of sum of independent random variables, notably its "distance" to the standard normal distribution play an important role in the classical central limit theorem.

A quite natural generalization to the classical central limit theorem is to study the asymptotic behavior of sum of independent random variables which are not necessary of Bernoulli's type (eventually not identically distributed). More precisely, for any integer $n \geq 1$ let $(X_{n,i})_{1\leq i\leq n}$ be a collection of independent random variables, and let $W_n = X_{n,1} + \cdots, X_{n,n}$. We want to study the convergence (in law) of the sequence $(W_n)_{n\geq 1}$ and the limit distribution if we have the convergence. The possible limit distribution of the sequence $(W_n)_{n\geq 1}$ must be an infinitely divisible distribution. Criteria have been given to the convergence of $(W_n)_{n\geq 1}$ to some infinitely divisible laws such as Normal laws, Poisson laws, or Dirac distributions. Interested reader can refer to Petrov [66] for a detailed review.

The convergence speed of central limit theorems has been largely studied. In general, the speed of convergence may be arbitrary slow. However, if we suppose the existence of certain order moments of $X_{n,k}$, we have more precise estimation of the convergence speed. The Berry-Esseen inequality states as follows.

Theorem 3.1.1 *(Berry-Esseen) Let $X_1, \cdots, X_n$ be independent zero-mean random variables having third order moment, $W = X_1 + \cdots + X_n$. Denote by $\sigma_i^2 = \mathrm{Var}(X_i)$, $\sigma_W^2 = \mathrm{Var}(W) = \sum_{i=1}^n \sigma_i^2$ and $F(x) = \mathbb{P}\Big(\frac{W}{\sigma_W} \leq x\Big)$, then*

$$\sup_{x\in\mathbb{R}} |F(x) - \mathcal{N}(x)| \leq \frac{A}{\sigma_W^3} \sum_{i=1}^n \mathbb{E}|X_i|^3,$$

where $\mathcal{N}$ is the distribution function of standard normal distribution, A is an absolute constant.

Theorem 3.1.1 gives a uniform upper bound of normal approximation error for the distribution function. For an arbitrary function h, the approximation error can be estimated by the Lindeberg method which consists of comparing two sum of independent random variables. More precisely, let $W = X_1 + \cdots + X_n$ and $S = \xi_1 + \cdots + \xi_n$ be two sums of independent random variables. If we write for any $0 \leq k \leq n$

$$U_k = \sum_{i=1}^k \xi_i + \sum_{j=k+1}^n X_j, \tag{3.5}$$

then we have $U_0 = W$ and $U_n = S$, and the difference $h(W) - h(S)$ is written as $\sum_{k=1}^n h(U_{k-1}) - h(U_k)$. Therefore

$$|\mathbb{E}[h(W)] - \mathbb{E}[h(S)]| \leq \sum_{k=1}^n \Big|\mathbb{E}[h(U_{k-1})] - \mathbb{E}[h(U_k)]\Big|.$$

Since U_k and U_{k-1} only differ by $\xi_k - X_k$, it is easier to estimate $\Big|\mathbb{E}[h(U_{k-1})] - \mathbb{E}[h(U_k)]\Big|$.

3.1.3 Stein's method and zero bias transformation

The *Stein's method* was first introduced by Stein [76] in 1972 to study the convergence rate of CLT for the standard normal distribution. Chen [15] extended it to the Poisson approximation. The method has then been developed by many authors and it provides a powerful tool for normal, Poisson and other approximations, in one and high dimensional cases, for independent or dependent random variables, or even for stochastic processes. The basic approach has been introduced in the monograph of Stein [77] himself. One may also consult Raic [68], Chen [16] or Chen and Shao [18] for a more detailed review. The Poisson approximation is comprehensively introduced in Barbour, Holst and Janson [7].

The *zero biasing*, or the *zero bias transformation*, is introduced by Goldstein and Reinert [39] in the framework of the Stein's method. In [39], the authors use the technique of zero bias transformation on functions satisfying certain moment conditions to derive the bounds of the approximation error. Some further development has been carried out by Goldstein [38] and Goldstein and Reinert [40]. This approach has many interesting properties. In particular, it provides a concise presentation which largely simplifies the writings and calculations. However, we surprisingly find that the usage of this method remains limited in the literature.

In the following, we begin our presentation of the zero bias transformation in Section 3.2 and we then discuss the Stein's method in Section 3.3.

3.2 Zero bias transformation and Gaussian distribution

In this section, we introduce the zero bias transformation and we present some estimation results in the normal approximation context. Two main results of the section are

1) Proposition 3.2.6 which enables us to calculate the expectation of functions on the difference between one random variable and its zero bias transformation when they are independent with an exact formula;

2) Proposition 3.2.16 which gives the estimations of the product of two functions where the variables are not independent. Instead of decomposing the sum variable into two independent parts, we use conditional expectations to estimate a covariance function and we obtain error bounds of one order higher than doing the estimation directly. This is the key argument we shall use in the following.

3.2.1 Definition and some known properties

The zero-bias transformation associated with a zero-mean, square integrable random variable is given as follows. In the following of this chapter, the symbol Z refers to a central normal variable, while X denotes a general central random variable.

Definition 3.2.1 (Goldstein and Reinert) Let X be a random variable with zero expectation and finite non-zero variance $\mathrm{Var}(X) = \sigma^2$. We say that a random variable X^* has the X-*zero biased distribution*, or that X^* is a *zero bias transformation* of X, if for any function f of C^1-type, whose derivative has compact support, we have

$$\mathbb{E}[Xf(X)] = \sigma^2 \mathbb{E}[f'(X^*)]. \tag{3.6}$$

The basic idea is based on the observation made by Stein in 1972 on the normal distribution: a random variable Z has the central normal distribution $N(0,\sigma^2)$ if and only if for any absolutely continuous function f such that $\mathbb{E}[f'(Z)]$ is well defined, we have

$$\mathbb{E}[Zf(Z)] = \sigma^2\mathbb{E}[f'(Z)]. \tag{3.7}$$

We observe that the zero biased distribution of the central normal distribution is itself. Therefore, it is natural to measure the distance between an arbitrary distribution and the central normal distribution by the distance between the given distribution and its zero biased distribution. If it's close to the normal distribution, then it should also be close to its zero biased distribution.

Another similar notion which inspired the zero bias transformation is the *size bias transformation* for nonnegative variables, which is defined, for any random variable $X \geq 0$ with $\mathbb{E}[X] = \mu < \infty$ and any function f such that $\mathbb{E}[Xf(X)]$ exists, by $\mathbb{E}[Xf(X)] = \mu\mathbb{E}[f(X^{\text{size}})]$. We say also that X^{size} has the *X-size biased distribution.* This notion and its relation between the Stein's method are discussed in Stein [78] and Goldstein and Rinott [41]. There are many similitudes between these two notions. However, as we are interested in the closeness of one distribution with the normal distribution, it's unnatural that we exclude symmetric random variables since X is required to be positive here. Hence it's more practical to work directly with the zero mean random variable by the zero bias transformation.

The existence of a random variable with zero bias distribution is given in [39] by providing the density function. In addition, as mentioned above, the Z-zero biased distribution associated with a random variable Z of the zero-mean normal distribution $N(0,\sigma^2)$ is the normal distribution $N(0,\sigma^2)$ itself. We here give the proof of the converse property.

Theorem 3.2.2 *(Goldstein and Reinert) Let X be a zero-mean random variable with finite variance $\sigma^2 > 0$.*

1) A random variable X^* with the following density $p_{X^*}(x)$ with respect to the Lebesgue measure has X-zero biased distribution.

$$p_{X^*}(x) = \sigma^{-2}\mathbb{E}[X\mathbb{1}_{\{X>x\}}]. \tag{3.8}$$

2) If Z and Z^* have the same distribution, then Z is a centered Gaussian variable.

Proof. We proceed by verification after having established the identity (3.8).

i) It is obvious that $p_{X^*} \geq 0$ if $x \geq 0$. If x is negative, using the assumption $\mathbb{E}[X] = 0$, we rewrite p_{X^*} as

$$p_{X^*}(x) = \sigma^{-2}\mathbb{E}[X\mathbb{1}_{\{X>x\}} - X] = \sigma^{-2}\mathbb{E}[-X\mathbb{1}_{\{-X\geq -x\}}] \geq 0.$$

ii) Let g be a bounded Borel function with compact support, and $G(x) = \int_{-\infty}^{x} g(t)dt$ be a primitive function of g. Then G is bounded, and $XG(X)$ is integrable. On the other hand, we have by Fubini's theorem

$$\begin{aligned} \sigma^{-2}\int_{\mathbb{R}} g(x)\mathbb{E}[X1\!\!1_{\{X>x\}}]dx &= \sigma^{-2}\mathbb{E}\Big[\int_{\mathbb{R}} g(x)X1\!\!1_{\{X>x\}}dx\Big] \\ &= \sigma^{-2}\mathbb{E}[XG(X)] = \varphi(g). \end{aligned} \tag{3.9}$$

Since φ is a positive fonctional on the space of continuous functions of compact support, by Riesz's theorem, it's a Radon measure on $\mathbb{R}$. Moreover, (3.9) means that φ has density $\sigma^{-2}\mathbb{E}[X1\!\!1_{\{X>x\}}]$. Finally, we verify that $\varphi(\mathbb{R}) = \sigma^{-2}\mathbb{E}[X^2] = 1$. Hence φ is a probability measure.

iii) If a random variable Z has the same distribution of Z^*, then Z admits a density function p, and this density function satisfies

$$p(x) = \sigma^{-2}\int_{x}^{\infty} t\,p(t)\,dt, \quad \text{or} \quad x\,p(x) - \sigma^2 p'(x) = 0.$$

The solutions of this differential equation are proportional (up to a constant) to $\exp\Big(-\dfrac{x^2}{2\sigma^2}\Big)$.

□

Remark 3.2.3 Note that the equality (3.6) is valid for a larger set of functions f. In fact, if suffices that f is an absolutely continuous function such that $\mathbb{E}[f'(X^*)]$ is well defined. Then in (3.9), $\mathbb{E}\big[\int_{\mathbb{R}} |f'(x)X1\!\!1_{\{X>x\}}|dx\big] < \infty$. By the Fubini's theorem, we obtain equation (3.6).

The following example is fundamental in what follows. It studies the Bernoulli random variable of zero mean and its zero bias transformation. Note here that we do not work directly with the standard Bernoulli variable of default indicator, but a normalized random variable taking two real values different from 0 and 1. In fact, for any random variable, we can apply the transformation to the centered variable $X - \mathbb{E}[X]$. This so-called asymmetric Bernoulli random variable satisfies the zero mean condition in the zero bias transformation and its two possible values are one positive and one negative since the expectation equals zero.

Example 3.2.4 (**Asymmetric Bernoulli**) Let X be a zero-mean asymmetric Bernoulli random variable taking two values $\alpha = q = 1-p$ and $\beta = -p$, $(0 < p, q < 1)$ in $[-1, 1]$, with probabilities $\mathbb{P}(X = q) = p$ and $\mathbb{P}(X = -p) = q = 1-p$ respectively. Then the first two moments of X are

$$\mathbb{E}(X) = 0, \quad \text{and} \quad \mathrm{Var}(X) = p\,q^2 + q\,p^2 = pq.$$

We denote this distribution by $\mathcal{B}(q,-p)$. Moreover, for any differentiable function f,

$$\frac{1}{\sigma^2}\mathbb{E}[Xf(X)] = \frac{1}{p\,q}\Big(p\,qf(q) - q\,pf(-p)\Big) = f(q) - f(-p) = \int_{-p}^{q} f'(t)dt,$$

which implies by Definition 3.2.1 that the zero bias distribution exists and is the uniform distribution on $[-p,q]$.
More generally, any zero-mean asymmetric Bernoulli random variable can be written as a dilatation of $\mathcal{B}(q,-p)$ by letting $\alpha = \gamma\,q$ and $\beta = -\gamma\,p$, which we denote by $\mathcal{B}_\gamma(q,-p)$. If X follows $\mathcal{B}_\gamma(q,-p)$, then $\mathrm{Var}(X) = \gamma^2\,pq$ and its X-zero bias distribution is the uniform distribution on $[-\gamma\,p,\gamma q]$.

3.2.2 Properties and estimations

In this subsection, we shall present some useful results of the zero bias transformation. Let X be a zero-mean square integrable random variable with finite variance $\sigma^2 > 0$ and X^* be a random variable having the X-zero biased distribution and independent to X. We are particularly interested in the estimation of functions on $|X - X^*|$, the quantity which is important in the normal approximation. Proposition 3.2.6 is based on the fact that X^* and X are independent.

Proposition 3.2.5 *If X has $(k+2)^{th}$-order moments, then X^* has k^{th}-order moments. Furthermore, we have the following equalities*

$$\mathbb{E}[|X^*|^k] = \frac{1}{\sigma^2}\frac{\mathbb{E}[|X|^{k+2}]}{k+1}, \quad \mathbb{E}[(X^*)^k] = \frac{1}{\sigma^2}\frac{\mathbb{E}[X^{k+2}]}{k+1}. \tag{3.10}$$

Proof. Let

$$F(x) = \frac{1}{k+1}|x|^k x,$$

then its derivative function $F'(x) = |x|^k$. If $\mathbb{E}[|X|^{k+2}]$ exists, then $\mathbb{E}[XF(X)]$ is well defined, and so is $\mathbb{E}[F'(X^*)]$. By definition, we have $\mathbb{E}[|X^*|^k] = \frac{1}{\sigma^2}\mathbb{E}[XF(X)] = \frac{1}{\sigma^2(k+1)}\mathbb{E}[|X|^{k+2}]$. For the same reason, the second equality also holds. $\square$

We shall often encounter, in the following, the calculation concerning the difference $X - X^*$. The estimations are easy when X and X^* are independent by using a symmetrical term $X^s = X - \widetilde{X}$, where $\widetilde{X}$ is an independent duplicate of X.

Proposition 3.2.6 *Let X and X^* be a pair of independent random variables, such that X^* has the X-biased distribution. Let g be a locally integrable even function and G be its primitive function defined by $G(x) = \int_0^x g(t)dt$. Then*

$$\mathbb{E}\big[g(X^* - X)\big] = \frac{1}{2\sigma^2}\mathbb{E}\big[X^s G(X^s)\big] \tag{3.11}$$

In particular,

$$\mathbb{E}[|X^* - X|] = \frac{1}{4\sigma^2}\mathbb{E}\big[|X^s|^3\big], \quad \mathbb{E}[|X^* - X|^k] = \frac{1}{2(k+1)\sigma^2}\mathbb{E}\big[|X^s|^{k+2}\big]. \tag{3.12}$$

Proof. By definition, for any real number K, we have

$$\sigma^2\mathbb{E}[g(X^* - K)] = \mathbb{E}[XG(X-K)].$$

Since X^* is independent of X, let $\widetilde{X}$ be a random variable having the same distribution and independent of X, then

$$\mathbb{E}[g(X^* - X)] = \frac{1}{\sigma^2}\mathbb{E}[\widetilde{X}G(\widetilde{X} - X)].$$

G is an odd function as g is even, then

$$\mathbb{E}[\widetilde{X}G(\widetilde{X} - X)] = \mathbb{E}[XG(X - \widetilde{X})] = -\mathbb{E}[XG(\widetilde{X} - X)],$$

which follows (3.11). To obtain (3.12), it suffices to let $g(x) = |x|$ and $g(x) = |x|^k$ respectively.

□

If X is a zero-mean asymmetric Bernoulli random variable which follows $\mathcal{B}_\gamma(q, -p)$ as in Example 3.2.4, that is $X = \gamma q$ with probability p and $X = -\gamma p$ with probability $q = 1-p$, the symmetrized random variable X^s takes the values 0 with probability $p^2 + q^2 = 1 - 2pq$, and the values γ or $-\gamma$ with probability $p\,q$. Thus we have

$$\mathbb{E}[|X^* - X|^k] = \frac{1}{2\gamma^2 pq}\frac{1}{k+1}|\gamma|^{k+2}2p\,q = \frac{1}{k+1}|\gamma|^k.$$

Remark 3.2.7 Equation (3.12) enables us to obtain an equality which is very useful in the estimation of error bounds. For example, in [39], $\mathbb{E}[|X - X^*|]$ is bounded by $\mathbb{E}[|X| + |X^*|]$. Our result enables to obtain a sharper bound.

Similar calculation yields estimates for the $\mathbb{P}(|X - X^*| \le \epsilon)$, giving a measure of the spread between X and X^*.

Corollary 3.2.8 *Let X and X^* be independent variables satisfying the conditions of Proposition 3.2.6. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}(|X - X^*| \le \varepsilon) \le \frac{\varepsilon}{\sqrt{2}\,\sigma} \wedge 1, \quad \mathbb{P}(|X - X^*| \ge \varepsilon) \le \frac{1}{4\sigma^2\,\varepsilon}\mathbb{E}\big[|X^s|^3\big] \tag{3.13}$$

Proof. Let us observe that the second inequality is immediate from the classical Markov inequality

$$\mathbb{P}(|X - X^*| \ge \varepsilon) \le \frac{1}{\varepsilon}\mathbb{E}\big[|X - X^*|\big].$$

To obtain the first inequality, we apply Proposition 3.2.6 to the even function $g(x) = \mathbb{1}_{\{|x|\leq\varepsilon\}}$ and its primitive $G(x) = \text{sign}(x)\big(|x| \wedge \varepsilon\big)$. So,

$$\mathbb{P}(|X - X^*| \leq \varepsilon) = \frac{1}{2\sigma^2}\mathbb{E}\big[|X^s|\big(|X^s| \wedge \varepsilon\big)\big] \tag{3.14}$$

Since $|X^s| \wedge \varepsilon \leq \varepsilon$ and $\mathbb{E}\big[|X^s|\big]^2 \leq \mathbb{E}\big[|X^s|^2\big] = 2\sigma^2$, we get

$$\mathbb{P}(|X - X^*| \leq \varepsilon) \leq \frac{\varepsilon}{2\sigma^2}(2\sigma^2)^{1/2} = \frac{\varepsilon}{\sqrt{2}\,\sigma}$$

□

Remark 3.2.9 The first inequality of Corollary 3.2.8 makes sense when ε is small, otherwise, the probability is always bounded by 1.

In particular, if X follows $\mathcal{B}_\gamma(q, -p)$, then we can calculate $\mathbb{P}(|X - X^*| \leq \varepsilon)$ explicitly. In fact, we have by (3.14)

$$\mathbb{P}(|X - X^*| \leq \varepsilon) = \frac{1}{2\gamma^2 pq} 2pq|\gamma|(|\gamma| \wedge \varepsilon) = \frac{|\gamma| \wedge \varepsilon}{|\gamma|}.$$

3.2.3 Sum of independent random variables

A typical example which concerns the sum of independent random variables deserves special attention. In fact, this example has been largely discussed in the Stein's method framework. The problem is relatively simple when we restrict to the most classical version where all variables are identically distributed. However, this elementary case can be extended to non-identically distributed variables. Goldstein and Reinert [39] give an interesting construction of zero bias transformation for the sum variable $W = X_1 + \cdots + X_n$ by replacing one single summand by its independent zero bias transformation variable. Such construction is informative since W and W^* differs only slightly. We now introduce the construction of zero biased distribution as in Goldstein and Reinert [39] for the sum of several independent non-identically distributed variables.

Proposition 3.2.10 *(Goldstein and Reinert) Let X_i $(i = 1, \cdots, n)$ be independent zero-mean random variables of finite variance $\sigma_i^2 > 0$ and X_i^* be random variables of the X_i-zero biased distribution. Denote by $(\overrightarrow{X}, \overrightarrow{X}^*) = (X_1, \cdots, X_n, X_1^*, \cdots, X_n^*)$ which are independent random variables.*
Let $W = X_1 + \cdots + X_n$ be the sum variable, and $\sigma_W^2 = \sigma_1^2 + \cdots + \sigma_n^2$ be its variance. We also use the notation $W^{(i)} = W - X_i$.
Let us introduce a random choice I of the index i such that $\mathbb{P}(I = i) = \sigma_i^2/\sigma_W^2$, and assume I independent of $(\overrightarrow{X}, \overrightarrow{X}^*)$.
Then the random variable $W^* = W^{(I)} + X_I^*$ has the W-zero biased distribution.

Proof. Let f be a continuous function with compact support and F be a primitive function of f. Then,

$$\mathbb{E}[WF(W)] = \sum_{i=1}^{n} \mathbb{E}[X_i F(W)]$$
$$= \sum_{i=1}^{n} \mathbb{E}[X_i F(W^{(i)} + X_i)] = \sum_{i=1}^{n} \sigma_i^2 \mathbb{E}[f(W^{(i)} + X_i^*)]$$

since X_i is independent of $W^{(i)}$. On the other hand, since I is independent of W,

$$\sigma_W^2 \mathbb{E}[f(W^{(I)} + X_I^*)] = \sum_{i=1}^{n} \sigma_i^2 \mathbb{E}[f(W^{(i)} + X_i^*)].$$

By comparing the above two equations, we know that $W^* = W^{(I)} + X_I^*$ has the W-zero biased distribution. □

Remark 3.2.11

1. From the above construction, W^* has the W-zero biased distribution, but is not independent of W. However the difference $W - W^* = X_I - X_I^*$ is easy to study, since X_I^* and X_I are conditionally independent given I.

2. If X_i are identically distributed, the probability of choosing a certain variate for the random index I is equal to $1/n$. Therefore, let $I = 1$, then $W^* = W^{(1)} + X_1^*$ has the W-zero biased distribution. However, for technical reasons, we insist on the usage of the random index representation $W^* = W^{(I)} + X_I^*$ where the random variable $X_I^* = \frac{1}{n}\sum_{i=1}^{n} X_i^*$ follows the same law with X_1^*.

3. In particular, if $X_1, \cdots, X_n$ are i.i.d. zero-mean asymmetric Bernoulli random variable which follow $\mathcal{B}_\gamma(q, -p)$, the sum W follows an asymmetric binomial distribution. Let the variance σ_W^2 of W be fixed, then the dilatation parameter is given by $\gamma = \frac{\sigma_W}{\sqrt{npq}}$.

The above proposition facilitates our study of the sum variable through the individual summand variables. We now extend the estimation results in the last subsection to the sum variable. In brackets, we show the moment order in the asymmetric binomial case.

Corollary 3.2.12 *With the notation of Proposition 3.2.10, we have*

$$\mathbb{E}[X_I^*] = \frac{1}{2\sigma_W^2} \sum_{i=1}^{n} \mathbb{E}[X_i^3] \quad \Big(\sim O\big(\frac{1}{\sqrt{n}}\big)\Big),$$

and

$$\mathbb{E}[(X_I^*)^2] = \frac{1}{3\sigma_W^2}\sum_{i=1}^n \mathbb{E}[X_i^4] \quad \Big(\sim O(\frac{1}{n})\Big)$$

and the following estimations

$$\mathbb{E}\big[|W^*-W|\big] = \frac{1}{4\sigma_W^2}\sum_{i=1}^n \mathbb{E}[|X_i^s|^3], \quad \mathbb{E}\big[|W^*-W|^k\big] = \frac{1}{2(k+1)\sigma_W^2}\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^{k+2}\big]. \tag{3.15}$$

In particular, for the asymmetric binomial case, we have

$$\mathbb{E}\big[|W^*-W|^k\big] = \frac{1}{k+1}\Big(\frac{\sigma_W}{\sqrt{np(1-p)}}\Big)^k. \tag{3.16}$$

Proof. In fact, the above results are obvious by using the definition of the zero bias transformation and the construction of W^*, together with previous estimations. □

We have in addition the estimation of the probability terms from the Corollary 3.2.8.

Corollary 3.2.13 *For any postive constant ε, we have*

$$\mathbb{P}(|W^*-W|\leq\varepsilon) \leq \Big(\frac{\varepsilon}{\sqrt{2}\sigma_W^2}\sum_{i=1}^n \sigma_i\Big)\wedge 1, \quad \mathbb{P}(|W^*-W|\geq\varepsilon) \leq \frac{1}{4\sigma_W^2\varepsilon}\sum_{i=1}^n \mathbb{E}[|X_i^s|^3]. \tag{3.17}$$

Proof. Proposition 3.2.10 and Corollary 3.2.8 imply immediately (3.17). □

Notice that X_I and X_I^* are not independent of W and W^*. However, we know the conditional expectation of X_I and X_I^* given $(\overrightarrow{X}, \overrightarrow{X}^*)$. This observation enables us to obtain some useful estimations which shall play an important role in the following.

Proposition 3.2.14 *We have the conditional expectation given by*

$$\mathbb{E}[X_I|X_1,\cdots,X_n] = \sum_{i=1}^n \frac{\sigma_i^2}{\sigma_W^2}X_i, \quad \mathbb{E}[\mathbb{E}[X_I|X_1,\cdots,X_n]^2] = \sum_{i=1}^n \frac{\sigma_i^6}{\sigma_W^4}. \tag{3.18}$$

Remark 3.2.15 We note that $\mathbb{E}[\mathbb{E}[X_I|X_1,\cdots,X_n]^2]$ is of order $O\big(\frac{1}{n^2}\big)$, which is significantly smaller than $\mathbb{E}[X_I^2]$ which is of order $O\big(\frac{1}{n}\big)$. In the homogenous case, if we take $W = W^{(1)} + X_1^*$, the above property no longer holds since $\mathbb{E}[X_1^2] \sim O(\frac{1}{n})$. This fact justifies the efficiency of the random index construction of W^*.

Proposition 3.2.16 *Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be two functions such that the variance of $f(W)$ exists, and that for all $i = 1, \cdots, n$, the variance of $g(X_i, X_i^*)$ exist, then*

$$\begin{aligned}&\big|\mathbb{E}[f(W)g(X_I, X_I^*)] - \mathbb{E}[f(W)]\mathbb{E}[g(X_I, X_I^*)]\big| \\ &\leq \frac{1}{\sigma_W^2}\mathrm{Var}[f(W)]^{\frac{1}{2}}\Big(\sum_{i=1}^n \sigma_i^4 \mathrm{Var}[g(X_i, X_i^*)]\Big)^{\frac{1}{2}}.\end{aligned} \tag{3.19}$$

Proof. We first notice that $\mathbb{E}[f(W)g(X_I, X_I^*)] = \mathbb{E}\big[f(W)\mathbb{E}[g(X_I, X_I^*)|\overrightarrow{X}, \overrightarrow{X}^*]\big]$ since W is the sum of $X_1, \cdots, X_n$. Therefore,

$$\mathbb{E}[f(W)g(X_I, X_I^*)] = \mathbb{E}[f(W)]\mathbb{E}[g(X_I, X_I^*)] + \mathrm{cov}(f(W), \mathbb{E}[g(X_I, X_I^*)|\overrightarrow{X}, \overrightarrow{X}^*]).$$

On the other hand, since (X_i, X_i^*) are mutually independent, we have

$$\begin{aligned}&\mathrm{cov}(f(W), \mathbb{E}[g(X_I, X_I^*)|\overrightarrow{X}, \overrightarrow{X}^*]) \\ &\leq \mathrm{Var}[f(W)]^{\frac{1}{2}}\mathrm{Var}\big[\mathbb{E}[g(X_I, X_I^*)|\overrightarrow{X}, \overrightarrow{X}^*]\big]^{\frac{1}{2}} \\ &\leq \frac{1}{\sigma_W^2}\mathrm{Var}[f(W)]^{\frac{1}{2}}\Big(\sum_{i=1}^n \sigma_i^4 \mathrm{Var}[g(X_i, X_i^*)]\Big)^{\frac{1}{2}}.\end{aligned}$$

□

Remark 3.2.17 Similar as in Remark 3.2.15, we here obtain the estimation of one order higher by using the conditional expectation than applying directly the Cauchy-Schwarz inequality. This is one of the key points in the estimations afterwards. The result holds when we replace W by W^* in (3.19).

We now apply the above proposition to obtain a useful estimation.

Corollary 3.2.18 *For any $\varepsilon > 0$, we have*

$$\big|\mathrm{cov}\big(\mathbb{1}_{\{a \leq W < b\}}, \mathbb{1}_{\{|X_I^* - X_I| \leq \varepsilon\}}\big)\big| \leq \frac{1}{2\sigma_W^2}\Big(\sum_{i=1}^n \frac{\sigma_i}{4\sqrt{2}}\mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}. \tag{3.20}$$

Proof. By Proposition 3.2.16, we get by using the conditional expectation that

$$\big|\mathrm{cov}\big(\mathbb{1}_{\{a \leq W < b\}}, \mathbb{1}_{\{|X_I^* - X_I| \leq \varepsilon\}}\big)\big| \leq \frac{1}{\sigma_W^2}\mathrm{Var}[\mathbb{1}_{\{a \leq W \leq b\}}]^{\frac{1}{2}}\Big(\sum_{i=1}^n \sigma_i^4 \mathrm{Var}[\mathbb{1}_{\{|X_i^* - X_i| \leq \varepsilon\}}]\Big)^{\frac{1}{2}}.$$

We have

$$\mathrm{Var}[\mathbb{1}_{\{a \leq W \leq b\}}] = \mathbb{P}(a \leq W \leq b)(1 - \mathbb{P}(a \leq W \leq b)) \leq \frac{1}{4}.$$

And we use Corollary 3.2.8 to get

$$\mathrm{Var}[\mathbb{1}_{\{|X_i^*-X_i|\leq\varepsilon\}}] \leq \frac{1}{4\sqrt{2}\sigma_i^3}\mathbb{E}[|X_i^s|^3],$$

which follows (3.20). □

Remark 3.2.19 Notice here the error bound of (3.20) does not depend on the value of ε. This property is useful when we prove the concentration inequality in Proposition 3.4.4.

3.3 Stein's equation

3.3.1 A brief review

We recall briefly the framework of the Stein's method. Consider a zero-mean random variable W of finite variance $\sigma_W^2 > 0$, which is the sum of n independent variables. Let $Z \sim N(0, \sigma_W^2)$. We are interested in the error of the normal approximation $\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]$ where h is some given function. Denote by $\Phi_{\sigma_W}(h) = \mathbb{E}[h(Z)]$. The Stein's method consists of associating this difference term with some auxiliary function by

$$\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h) = \mathbb{E}[Wf(W) - \sigma_W^2 f'(W)] \tag{3.21}$$

where f is the solution of the *Stein's equation* defined as the following differential equation

$$xf(x) - \sigma^2 f'(x) = h(x) - \Phi_\sigma(h). \tag{3.22}$$

Stein [77] studied some properties of the function f and gave estimations of $|f|$, $|xf|$ and $|f'|$ for the indicator function $h(x) = \mathbb{1}_{\{\tau\leq t\}}$. He mainly used inequalities of the Gaussian functions. We here need to consider the case when h is the call function.

The connection between the zero bias transformation with the Stein's method is evident by its definition. The difference between two expectations of a given function h for a zero-mean variable W and one central normal variable Z can be written as

$$\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h) = \mathbb{E}[Wf(W) - \sigma_W^2 f'(W)] = \sigma_W^2\mathbb{E}[f'(W^*) - f'(W)]. \tag{3.23}$$

Therefore, for the normal approximation, it is equivalent to study $\mathbb{E}[f'(W^*) - f'(W)]$, which is the difference of two expectations of the same function f' on W and on W^*. Recall $W^* = W^{(I)} + X_I^*$ in Proposition 3.2.10, we have intuition that W should not be "far" from W^*. In fact, compared to the Lindeberg method in which we change the summand variables successively, this method consists of changing the variable X_I to X_I^*.

We have discussed in the previous section the estimations concerning $|W - W^*|$. In the following of this section, we concentrate on the estimation concerning the function f. We propose two methods in Subsection 3.3.2 and Subsection 3.3.3 respectively. The first one is to extend the method used by Stein. Since the derivative of the call function is an indicator function, the techniques are similar. The second method consists of rewriting, by the Stein's equation, the auxiliary function as that of another function which is of slower growing speed. The method is efficient for polynomially growing functions and can be adapted to estimate high order expansions in Chapter 4.

We now give some properties of the solution of (3.22).

Proposition 3.3.1 *(Stein) If $h(t)\exp(-\frac{t^2}{2\sigma^2})$ is integrable on $\mathbb{R}$, then one solution of (3.22) is given by*

$$f(x) = \frac{1}{\sigma^2\phi_\sigma(x)}\int_x^\infty (h(t) - \Phi_\sigma(h))\phi_\sigma(t)dt, \tag{3.24}$$

where $\phi_\sigma(x)$ is the density function of the normal distribution $\mathcal{N}(0,\sigma^2)$ or equivalently by

$$f(x) = \frac{\sqrt{2\pi}}{\sigma}\mathbb{E}\big[\bar{h}(Z+x)e^{-\frac{Zx}{\sigma^2}}1\!\!1_{\{Z>0\}}\big], \tag{3.25}$$

where $\bar{h}(t) = h(t) - \Phi_\sigma(h)$ and $Z \sim N(0,\sigma^2)$.

Proof. Multiplying by $\sigma^{-2}\exp(-\frac{x^2}{2\sigma^2})$ on the two sides of (3.22), it's easy to obtain a global solution

$$f(x) = \frac{1}{\sigma^2}\exp\big(\frac{x^2}{2\sigma^2}\big)\int_{-\infty}^x (\Phi_\sigma(h) - h(t))e^{-\frac{t^2}{2\sigma^2}}dt \tag{3.26}$$

when $h(t)\exp(-\frac{t^2}{2\sigma^2})$ is integrable. In addition, by definition, $\Phi_\sigma(h) = \int_{-\infty}^\infty h(t)\phi_\sigma(t)dt$, which implies (3.24). We write (3.24) as

$$f(x) = \frac{1}{\sigma^2}\exp\big(\frac{x^2}{2\sigma^2}\big)\int_x^\infty (h(t) - \Phi_\sigma(h))e^{-\frac{t^2}{2\sigma^2}}dt.$$

Then by a change of variable $u = t - x$, we get

$$f(x) = \frac{\sqrt{2\pi}}{\sigma}\int_0^\infty \bar{h}(u+x)e^{-\frac{ux}{\sigma^2}}\phi_\sigma(u)du,$$

which implies immediately (3.25). □

Hence, by replacing x with W and σ with σ_W in equation (3.22) and taking expectations on the two sides, we verify that f is the solution of (3.21). Furthermore, we denote by $\mathcal{N}_\sigma(x)$ the distribution function of $N(0,\sigma^2)$, then another alternative form of a solution is given by $f(x) = \frac{1}{\sigma^2\phi_\sigma(x)}\Big(\mathcal{N}_\sigma(x)\Phi_\sigma(h) - \mathbb{E}[h(Z)1\!\!1_{\{Z\le x\}}]\Big)$.

In the following, we shall denote by $f_{h,\sigma}$ the solution (3.24) of the equation (3.22). When there is no ambiguity, we write simply f_h instead of $f_{h,\sigma}$. Clearly f_h is linear on h. From the Proposition 3.3.1, the integral form (3.24) shows that the function f_h is once more differentiable than h. The equation (3.22) was first discussed by Stein for the case $\sigma = 1$. The expectation form is introduced in Barbour [3] with which he deduces some estimations for the derivatives. In the following, we shall use different methods to estimate the derivatives of f_h according to the two forms (3.24) and (3.25) respectively. Furthermore, comparing (3.24) and (3.26), we have the equality $f_h(-x) = -f_{h(-t)}(x)$ which will sometimes simplify the discussion.

3.3.2 Estimations with the expectation form

It is shown in the above that the normal approximation error is related to the auxiliary function f_h. Hence, we are interested in some bound estimations concerning the function f_h. In this subsection, we shall give estimations based on the expectation form of f_h where h is the indicator function and the call function. The method used here was presented in Stein [77]. The more general case was studied in Barbour [3] to get higher order estimations. In the next subsection, we propose a new method based on the integral form of f_h.

For any real number α let I_α be the indicator function $I_\alpha(x) = 1\!\!1_{\{x\leq\alpha\}}$, and let $C_\alpha = (x-\alpha)_+$ be the "Call" function. We first recall the inequality concerning the normal distribution functions, which can be found in Stein [77] and Chen and Shao [18].

Proposition 3.3.2 *Denote by $\phi_\sigma(x)$ the density function and $\mathcal{N}_\sigma(x)$ the cumulative distribution function of the central normal distribution $N(0,\sigma^2)$, then*

$$\begin{cases} 1-\mathcal{N}_\sigma(x) < \frac{\sigma^2\phi_\sigma(x)}{x}, & x>0 \\ \mathcal{N}_\sigma(x) < \frac{\sigma^2\phi_\sigma(x)}{|x|}, & x<0. \end{cases} \tag{3.27}$$

Proof. We first consider the case where $x > 0$. Notice that $\phi'_\sigma(x) = -\frac{x}{\sigma^2}\phi_\sigma(x)$, then direct calculation gives that

$$1-\mathcal{N}_\sigma(x) = -\int_x^\infty \frac{\sigma^2}{t} d\phi_\sigma(t) < \frac{\sigma^2}{x}\phi_\sigma(x).$$

The case where $x < 0$ is similar. □

For technical reason, we introduce the following notation: for any function h such that $h\phi_\sigma$ is integrable on $(-\infty,-x)\cup(x,+\infty)$ for any $x>0$, let $\widetilde{f}_{h,\sigma}$ be the function defined over $\mathbb{R}\setminus\{0\}$ by

$$\widetilde{f}_{h,\sigma}(x) = \begin{cases} \frac{1}{\sigma^2\phi_\sigma(x)}\int_x^\infty h(t)\phi_\sigma(t)dt, & x>0 \\ -\frac{1}{\sigma^2\phi_\sigma(x)}\int_{-\infty}^x h(t)\phi_\sigma(t)dt, & x<0. \end{cases} \tag{3.28}$$

We write $\widetilde{f}_h$ instead of $\widetilde{f}_{h,\sigma}$ when there is no ambiguity. We also give the expectation form of $\widetilde{f}_h$:

$$\widetilde{f}_h(x) = \begin{cases} \frac{\sqrt{2\pi}}{\sigma}\mathbb{E}\big[h(Z+x)e^{-\frac{Zx}{\sigma^2}}1\!\!1_{\{Z>0\}}\big], & x > 0 \\ \frac{\sqrt{2\pi}}{\sigma}\mathbb{E}\big[h(Z+x)e^{-\frac{Zx}{\sigma^2}}1\!\!1_{\{Z<0\}}\big], & x < 0 \end{cases} \tag{3.29}$$

where $Z \sim N(0, \sigma^2)$. Notice that in general, $\widetilde{f}_h$ can not be extended as a continuous function on $\mathbb{R}$. if $\mathbb{E}[|h(Z)|] < +\infty$, we have $\widetilde{f}_h(0-) = -\frac{\sqrt{2\pi}}{\sigma}\mathbb{E}[h(Z)1\!\!1_{\{Z<0\}}]$ and $\widetilde{f}_h(0+) = \frac{\sqrt{2\pi}}{\sigma}\mathbb{E}[h(Z)1\!\!1_{\{Z>0\}}]$. The two limits are equal if and only if $\Phi_\sigma(h) = \mathbb{E}[h(Z)] = 0$. Furthermore, if $\Phi_\sigma(h) = 0$, then $\widetilde{f}_h$ coincides with the solution f_h of the Stein's equation.

We introduce the definition set $\mathscr{E}$ of $\widetilde{f}_h$ as below and we only study $h \in \mathscr{E}$ in the following. Let $\mathscr{E}$ be the set of functions g defined on $\mathbb{R} \setminus \{0\}$ taking values on $\mathbb{R}$ such that g is locally of finite variation and has finite number of jump points and that g satisfies $\int |g(x)|\phi_\sigma(x)1\!\!1_{\{|x|>a\}}dx < \infty$ for any $a > 0$. It's evident that (3.28) is well defined for any $h \in \mathscr{E}$. In fact, the above condition specifies the regularity of functions we are interested in and we exclude the "irregular" functions which are not contained in $\mathscr{E}$.

Proposition 3.3.3 *We have following properties of $\widetilde{f}_h$ for any $x \in \mathbb{R} \setminus \{0\}$:*

1. $f_h(x) = \widetilde{f}_{\bar{h}}(x)$ where $\bar{h} = h - \Phi_\sigma(h)$;

2. $\widetilde{f}_h(-x) = -\widetilde{f}_{h(-t)}(x)$;

3. $\widetilde{f}_h$ is one solution of the following equation

$$x\widetilde{f}_h(x) - \sigma^2 \widetilde{f}_h'(x) = h(x). \tag{3.30}$$

Proof. 1) and 2) are directly by definition. For 3), it is easy to verify that $\widetilde{f}_h$ defined by (3.28) is one solution of the differential equation (3.30). □

Remark 3.3.4 We call (3.30) the *decentralized Stein's equation.* It is useful to introduce $\widetilde{f}_h$ since by taking derivatives, we shall often work with non-centralized functions. Moreover, by 1) of Proposition 3.3.3, the properties of f_h can be deduced directly.

We first give some simple properties of the function $\widetilde{f}_h$.

Proposition 3.3.5 *1) Let h_1 and h_2 be two functions and a_1 and a_2 be two real numbers. Then $\widetilde{f}_{a_1h_1+a_2h_2} = a_1\widetilde{f}_{h_1} + a_2\widetilde{f}_{h_2}$.*

2) If $|h(x)| \le g(x)$, then $|\widetilde{f}_h| \le |\widetilde{f}_g|$.

3) If $|h(x)| \leq g(x)$ and if $\frac{g(x)}{|x|}$ is decreasing when $x > 0$ and is increasing when $x < 0$, then $|\widetilde{f}_h(x)| \leq |\frac{g(x)}{x}|$.

Proof. 1) and 2) are evident by definition.

For 3), we first study the case where $x > 0$. By definition of $\widetilde{f}_h$, we know that $|\widetilde{f}_h(x)| \leq \frac{1}{\sigma^2 \phi_\sigma(x)} \int_x^\infty g(t)\phi_\sigma(t)dt$. Notice that $\phi'_\sigma(x) = -\frac{x}{\sigma^2}\phi_\sigma(x)$, then

$$|\widetilde{f}_h(x)| \leq -\frac{1}{\phi_\sigma(x)} \int_x^\infty \frac{g(t)}{t} d\phi_\sigma(t).$$

Since $\frac{g(x)}{x}$ is decreasing, we get the inequality. When $x < 0$, the proof is similar. □

Corollary 3.3.6 *We have $|x\widetilde{f}_1(x)| \leq 1$.*

Proof. It is a direct consequence of Proposition 3.3.5 applied to $h = g = 1$. □

The following two propositions allow us to estimate the derivatives of f_h in the expectation form. The argument is based on the fact that the polynomial functions increase slower than the exponential functions.

Proposition 3.3.7 *Let $Z \sim N(0, \sigma^2)$. Then for any non-negative integers l, m satisfying $l \leq m$ and for any $x > 0$,*

$$\mathbb{E}\big[\mathbb{1}_{\{Z>0\}} x^l Z^m e^{-\frac{Zx}{\sigma^2}}\big] \leq \frac{1}{2}\big(\frac{l\sigma^2}{e}\big)^l \mathbb{E}\big[|Z|^{m-l}\big]. \tag{3.31}$$

Where, by convention, $0^0 = 1$.

Proof. Consider the function $f(y) = y^l e^{-\frac{y}{\sigma^2}}$, it attains the maximum value at $y = l\sigma^2$, then $|f(y)| \leq (\frac{l\sigma^2}{e})^l$. Then the lemma follows immediately. □

Proposition 3.3.8 *Let $Z \sim N(0, \sigma^2)$. Then for any $x > 0$ we have*

$$x\mathbb{E}[\mathbb{1}_{\{Z>0\}} e^{-\frac{Zx}{\sigma^2}}] \leq \frac{\sigma}{\sqrt{2\pi}}. \tag{3.32}$$

Proof. If suffices to observe that $x\mathbb{E}[\mathbb{1}_{\{Z>0\}} e^{-\frac{Zx}{\sigma^2}}] = \frac{\sigma}{\sqrt{2\pi}} x\widetilde{f}_1(x)$. Then applying Corollary 3.3.6 gives (3.32). □

Remark 3.3.9 For the case $x < 0$ in (3.31) and (3.32), we can obtain by symmetry the following inequalities for any integers $0 \leq l \leq m$:

$$\mathbb{E}[\mathbb{1}_{\{Z<0\}} |x|^l |Z|^m e^{-\frac{Zx}{\sigma^2}}] \leq \frac{1}{2}\left(\frac{l\sigma^2}{e}\right)^l \mathbb{E}[|Z|^{m-l}]$$

and $\mathbb{E}[\mathbb{1}_{\{Z<0\}}|x|e^{-\frac{Zx}{\sigma^2}}] \leq \frac{\sigma}{\sqrt{2\pi}}$.

Corollary 3.3.10 *For any $x \in \mathbb{R}$, $|\widetilde{f}_1(x)| \leq \frac{\sqrt{2\pi}}{2\sigma}$.*

Proof. Since $\lim_{x\to 0+} \widetilde{f}_1(x) = \sqrt{2\pi}/2\sigma$, $\lim_{x\to 0-} \widetilde{f}_1(x) = -\sqrt{2\pi}/2\sigma$, and $\lim_{|x|\to+\infty} \widetilde{f}_1(x) = 0$, we only need to prove that $\widetilde{f}_1$ is decreasing when $x > 0$ and when $x < 0$ respectively. In fact, by Corollary 3.3.6, we have $\widetilde{f}_1'(x) = \frac{1}{\sigma^2}(x\widetilde{f}_1(x) - 1) \leq 0$ for any $x > 0$ and $\widetilde{f}_1'(x) = -\frac{1}{\sigma^2}(1 - x\widetilde{f}_1(x)) \leq 0$ for any $x < 0$. □

By Proposition 3.3.5 2), we can give the upper bound of $\widetilde{f}_h$ and $\widetilde{f}_h'$ for all bounded functions h by Corollary 3.3.6 and Corollary 3.3.10 as below.

Proposition 3.3.11 *Let h be a bounded function on $\mathbb{R}$ and let $c_0 = \|h\|$, then*

1) $|\widetilde{f}_h(x)| \leq \sqrt{2\pi}c_0/2\sigma$,

2) $|\widetilde{f}_h'(x)| \leq 2c_0/\sigma^2$,

Proof. 1) is direct by Proposition 3.3.5 since $|h(x)| \leq c_0$ for any $x \in \mathbb{R}$.

2) By Stein's equation, $\widetilde{f}_h'(x) = \frac{1}{\sigma^2}(x\widetilde{f}_h(x) - h(x))$. Notice that for any $x \in \mathbb{R}$,

$$|x\widetilde{f}_h(x)| \leq |c_0 x\widetilde{f}_1(x)| \leq c_0, \quad |h(x)| \leq c_0.$$

So $|\widetilde{f}_h'(x)| \leq 2c_0/\sigma^2$. □

The indicator function satisfies the boundedness condition with $c_0 = 1$. So Proposition 3.3.11 applies directly. Let $I_\alpha(x) = \mathbb{1}_{\{x\leq\alpha\}}$. We now give the estimation for xf_{I_α}'. For technique reasons, we consider a small lag of β where $0 \leq \beta \leq 1$. Then $0 \leq |I_\alpha - \beta| \leq 1$ and we shall see that the bound is uniform on β.

Proposition 3.3.12 *For any real number $\beta \in [0, 1]$,*

$$|x\widetilde{f}_{I_\alpha-\beta}'(x)| \leq \frac{\sqrt{2\pi}}{2\sigma e} + \frac{|\alpha|}{\sigma^2}.$$

Proof. First we consider the case $x > 0$, by definition,

$$\widetilde{f}_{I_\alpha-\beta}(x) = \frac{\sqrt{2\pi}}{\sigma}\mathbb{E}[\mathbb{1}_{\{Z>0\}}(I_\alpha(x+Z) - \beta)e^{-\frac{Zx}{\sigma^2}}].$$

Then

$$\widetilde{f}_{I_\alpha-\beta}'(x) = -\frac{\sqrt{2\pi}}{\sigma^3}\mathbb{E}[\mathbb{1}_{\{Z>0\}}(I_\alpha(x+Z) - \beta)Ze^{-\frac{Zx}{\sigma^2}}] + \mathbb{1}_{\{x\leq\alpha\}}\frac{1}{\sigma^2}e^{\frac{x^2-\alpha^2}{2\sigma^2}}.$$

Using Proposition 3.3.7 with $l = m = 1$ and the fact that $\|I_\alpha - \beta\| \leq 1$ we get

$$\left| x\mathbb{E}[\mathbb{1}_{\{Z>0\}}(I_\alpha(x+Z) - \beta)Ze^{-\frac{Zx}{\sigma^2}}] \right| \leq \frac{\sigma^2}{2e},$$

and

$$x\mathbb{1}_{\{x\leq\alpha\}}\frac{1}{\sigma^2}e^{\frac{x^2-\alpha^2}{2\sigma^2}} \leq \frac{|\alpha|}{\sigma^2}.$$

So combining the two terms, we have

$$|x\widetilde{f}'_{I_\alpha-\beta}(x)| \leq \frac{\sqrt{2\pi}}{2\sigma e} + \frac{|\alpha|}{\sigma^2}.$$

By a similar calculation for $x < 0$, we get the proposition. □

We can now resume the estimations of f_{I_α} by using $\widetilde{f}_{I_\alpha}$ for the indicator function.

Corollary 3.3.13 *Let $I_\alpha(x) = \mathbb{1}_{\{x\leq\alpha\}}$. Then*

$$\|f_{I_\alpha}\| \leq \frac{\sqrt{2\pi}}{2\sigma}, \quad \|f'_{I_\alpha}\| \leq \frac{2}{\sigma^2},$$

and

$$|xf'_{I_\alpha}| \leq \frac{\sqrt{2\pi}}{2\sigma e} + \frac{|\alpha|}{\sigma^2}.$$

Proof. By Proposition 3.3.3, $f_{I_\alpha} = \widetilde{f}_{\bar{I}_\alpha}$ where $\bar{I}_\alpha = I_\alpha - \mathbb{P}(Z \leq \alpha) = I_\alpha - \mathcal{N}_\sigma(\alpha)$. Since $|\bar{I}_\alpha| \leq 1$, we can apply Proposition 3.3.11 to obtain the first two inequalities. Proposition 3.3.12 implies the third one. □

We have presented above a quite formal way to estimate $\widetilde{f}_h$ and its derivatives and the method is easy to apply. However, the bound estimation we get is not always optimal. Stein [77] gave the following estimations for the indicator function:

$$0 < |f_{I_\alpha}(x)| \leq \min\Big(\frac{\sqrt{2\pi}}{4\sigma}, \frac{1}{|x|}\Big), \quad |f'_{I_\alpha}| \leq \frac{1}{\sigma^2}.$$

Compared to the constants obtained by Stein, those in Corollary 3.3.13 are twice larger.

In the following, we consider the functions with bounded derivatives. The increasing speed of these functions are at most linear. The call function satisfies this property.

Proposition 3.3.14 *Let h be an absolutely continuous function on $\mathbb{R}$.*

1) Let $c_1 = |h(0)|$ and suppose that $c_0 = \|h'\| < +\infty$, then

$$|\widetilde{f}_h(x)| \leq \frac{\sqrt{2\pi}c_1}{2\sigma} + 2c_0, \quad |\widetilde{f}'_h(x)| \leq \frac{\sqrt{2\pi}c_0}{\sigma}\Big(1 + \frac{1}{2e}\Big) + \frac{c_1}{\sigma^2}.$$

2) If, in addition, $c_2 = \|h\| < +\infty$, then

$$|x\widetilde{f}_h'(x)| \leq c_0 + \frac{\sqrt{2\pi}c_2}{2\sigma e}.$$

3) If, in addition to the hypotheses of 1), we assume that $h' \in \mathscr{E}$. Let $h' = g_1 + g_2$, where g_1 is the continuous part of h' and g_2 is the pure jump part of h' of the following form

$$g_2(x) = \sum_{i=1}^{N} \epsilon_i (I_{\mu_i} - \beta_i).$$

We assume that $c_3 = \|g_1'\| < +\infty$ and $c_4 = \|g_1\|$, then

$$|x\widetilde{f}_h''(x)| \leq c_3 + \frac{\sqrt{2\pi}c_4}{2\sigma e} + \sum_{i=1}^{N} |\varepsilon_i| \Big(\frac{\sqrt{2\pi}}{2\sigma e} + \frac{|\mu_i|}{\sigma^2} \Big) + \frac{1}{e\sigma} \Big(\frac{c_1}{\sigma} + \sqrt{2\pi}c_0 + \frac{2\sqrt{2\pi}c_0}{e} \Big).$$

Proof. Clearly we have $|h(x)| \leq c_1 + c_0|x|$ for any $x \in \mathbb{R}$. By a symmetric argument it suffices to prove the inequalities for $x > 0$.

1) As $|h(x)| \leq c_1 + c_0|x|$, we have

$$|\widetilde{f}_h(x)| \leq \frac{\sqrt{2\pi}}{\sigma} \mathbb{E}[1\!\!1_{\{Z>0\}}(c_1 + c_0 Z + c_0 x)e^{-\frac{Zx}{\sigma^2}}].$$

Then inequalities (3.31) and (3.32) yield

$$|\widetilde{f}_h(x)| \leq \frac{\sqrt{2\pi}}{\sigma} \Big[\frac{c_1}{2} + \frac{c_0}{2}\mathbb{E}[|Z|] + \frac{c_0\sigma}{\sqrt{2\pi}} \Big] \leq \frac{\sqrt{2\pi}c_1}{2\sigma} + 2c_0$$

since $\mathbb{E}[|Z|] = \frac{2\sigma}{\sqrt{2\pi}}$. Taking the derivative,

$$\widetilde{f}_h'(x) = \frac{\sqrt{2\pi}}{\sigma} \mathbb{E}[1\!\!1_{\{Z>0\}}h'(Z+x)e^{-\frac{Zx}{\sigma^2}}] - \frac{\sqrt{2\pi}}{\sigma^3} \mathbb{E}[1\!\!1_{\{Z>0\}}Zh(Z+x)e^{-\frac{Zx}{\sigma^2}}]. \tag{3.33}$$

Notice that the first term is $\widetilde{f}_{h'}(x)$ with h' being bounded and the second term can be estimated as above. So

$$\begin{aligned} |\widetilde{f}_h'(x)| &\leq \frac{\sqrt{2\pi}c_0}{2\sigma} + \frac{\sqrt{2\pi}}{\sigma^3} \mathbb{E}[1\!\!1_{\{Z>0\}}Z(c_0 Z + c_0 x + c_1)e^{-\frac{Zx}{\sigma^2}}] \\ &\leq \frac{\sqrt{2\pi}c_0}{2\sigma} + \frac{\sqrt{2\pi}}{\sigma^3} \Big(\frac{c_0\sigma^2}{2} + \frac{c_0\sigma^2}{2e} + \frac{c_1\sigma}{\sqrt{2\pi}} \Big) = \frac{\sqrt{2\pi}c_0}{\sigma} \Big(1 + \frac{1}{2e} \Big) + \frac{c_1}{\sigma^2}. \end{aligned}$$

2) If, in addition, h itself is bounded by c_2, we apply the above method by using (3.33) to get

$$|x\widetilde{f}_h'(x)| \leq c_0 + \frac{\sqrt{2\pi}c_2}{2\sigma e}.$$

3) By (3.33),

$$\widetilde{f}_h'' = \widetilde{f}_{h'}' - \frac{\sqrt{2\pi}}{\sigma^3}\mathbb{E}[\mathbb{1}_{\{Z>0\}} Z h'(Z+x) e^{-\frac{Zx}{\sigma^2}}] + \frac{\sqrt{2\pi}}{\sigma^5}\mathbb{E}[\mathbb{1}_{\{Z>0\}} Z^2 h(Z+x) e^{-\frac{Zx}{\sigma^2}}].$$

By the linearity of $\widetilde{f}_h$ with respect to h, we know that

$$\widetilde{f}_{h'}' = \widetilde{f}_{g_1}' + \sum_{i=1}^{N} \epsilon_i \widetilde{f}_{I_{\mu_i}-\beta_i}'.$$

So Proposition 3.3.12 and 2) imply that

$$|x\widetilde{f}_{h'}'(x)| \leq |x\widetilde{f}_{g_1}'(x)| + \sum_{i=1}^{N} |\epsilon_i| \Big(\frac{\sqrt{2\pi}}{2\sigma e} + \frac{|\mu_i|}{\sigma^2} \Big) \leq c_3 + \frac{\sqrt{2\pi} c_4}{2\sigma e} + \sum_{i=1}^{N} |\epsilon_i| \Big(\frac{\sqrt{2\pi}}{2\sigma e} + \frac{|\mu_i|}{\sigma^2} \Big).$$

The other two terms are estimated by (3.31) and (3.32) as above,

$$\Big| x\mathbb{E}[\mathbb{1}_{\{Z>0\}} Z h'(Z+x) e^{-\frac{Zx}{\sigma^2}}] \Big| \leq \frac{c_0\sigma^2}{2e}$$

$$\Big| x\mathbb{E}[\mathbb{1}_{\{Z>0\}} Z^2 h(Z+x) e^{-\frac{Zx}{\sigma^2}}] \Big| \leq \frac{c_0\sigma^4}{2e} + \frac{2c_0\sigma^4}{e^2} + \frac{c_1\sigma^3}{\sqrt{2\pi} e}.$$

So we get finally

$$|x\widetilde{f}_h''(x)| \leq c_3 + \frac{\sqrt{2\pi} c_4}{2\sigma e} + \sum_{i=1}^{N} |\varepsilon_i| \Big(\frac{\sqrt{2\pi}}{2\sigma e} + \frac{|\mu_i|}{\sigma^2} \Big) + \frac{1}{e\sigma} \Big(\frac{c_1}{\sigma} + \sqrt{2\pi} c_0 + \frac{2\sqrt{2\pi} c_0}{e} \Big).$$

□

For the call function, we apply directly the above Proposition.

Corollary 3.3.15 *Let $C_k = (x-k)^+$, then*

$$\|f_{C_k}\| \leq 2 + \frac{\sqrt{2\pi}}{2\sigma} c_1$$

where $c_1 = |(-k)^+ - \bar{c}|$ and $\bar{c} = \Phi_\sigma((x-k)^+) = \sigma^2 \phi_\sigma(k) - k(1 - \Phi_\sigma(k))$.

$$\|f_{C_k}'\| \leq \frac{\sqrt{2\pi}}{\sigma} \Big(1 + \frac{1}{2e} \Big) + \frac{c_1}{\sigma^2}$$

and

$$|x f_{C_k}''| \leq \frac{c_1}{e\sigma^2} + \frac{|k|}{\sigma^2} + \frac{2\sqrt{2\pi}}{\sigma e} \Big(1 + \frac{1}{e} \Big).$$

Proof. We have $f_{C_k} = \widetilde{f}_{\overline{C}_k}$ where $\overline{C}_k = C_k - \bar{c}$. In addition, $\|\overline{C}_k'\| = 1$ and $c_1 = |\overline{C}_k(0)| = |(-k)^+ - \bar{c}|$. Applying Proposition 3.3.14, we get the first inequalities. And if suffices to notice $c_3 = 0$ and $c_4 = 1$ to end the proof. □

3.3.3 Estimations with the integral form

In this subsection, we give another method to estimate the derivatives of f_h which is based on the integral form (3.24). The idea is to work with another function whose growing speed is one degree less than h and to deduce by a recurrence procedure.

The following lemma shows that to study $\widetilde{f}_h$, we can work with a more smooth function by the zero bias transformation.

Proposition 3.3.16 *Let h be a given function and H be one primitive function of h. Then*

$$\widetilde{f}_{xH}(x) = H + \sigma^2 \widetilde{f}_h(x). \tag{3.34}$$

Moreover, $\widetilde{f}'_{xH}(x) = x\widetilde{f}_h(x)$.

Proof. In fact, it's easy to verify that the right side of (3.34) is the solution of the equation $x\widetilde{f}_{xH}(x) - \sigma^2 \widetilde{f}'_{xH}(x) = xH(x)$. Then taking derivative gives immediately $\widetilde{f}'_{xH}(x) = h + \sigma^2 \widetilde{f}'_h(x) = x\widetilde{f}_h(x)$. □

Corollary 3.3.17 *Let h be a function such that $\mathbb{E}[h(W^*)]$ exists, then*

$$\mathbb{E}[h(W^*)] = \mathbb{E}\big[W^* \widetilde{f}_h(W^*) - W\widetilde{f}_h(W)\big].$$

Proof. Let H be a primitive function of h, then $\mathbb{E}[h(W^*)] = \frac{1}{\sigma_W^2}\mathbb{E}[WH(W)]$, which from the decentralized Stein's equation (3.30), equals $\frac{1}{\sigma_W^2}\mathbb{E}[W\widetilde{f}_{xH}(W) - \sigma_W^2 \widetilde{f}'_{xH}(W)]$. Then from the above Lemma, we have

$$\mathbb{E}[h(W^*)] = \mathbb{E}\big[\widetilde{f}'_{xH}(W^*) - \widetilde{f}'_{xH}(W)\big] = \mathbb{E}\big[W^* \widetilde{f}_h(W^*) - W\widetilde{f}_h(W)\big].$$

□

The following corollary gives a reverse version of Proposition 3.3.16 by letting $g = xH$. Then $h = \big(\frac{g(x)}{x}\big)'$. This writing facilitates the calculation and provides a useful method of estimation when x is not around zero. In the estimations afterwards, we shall distinguish this case. Usually we consider the cases when $|x| > 1$ and when $|x| \leq 1$ respectively.

Let g be an absolutely continuous function and we define the operator $\Gamma(g)$ for any $x \neq 0$ by

$$\Gamma(g) = \Big(\frac{g(x)}{x}\Big)'. \tag{3.35}$$

In the following, we suppose that $\Gamma(g) \in \mathscr{E}$, which means that g is a function such that $g' \in \mathscr{E}$ and that the function $\big|(\frac{g(x)}{x})'\big|\phi_\sigma(x)$ is integrable on $(-\infty, -a) \cup (a, +\infty)$ for any $a > 0$.

Corollary 3.3.18 *Let g be a function such that $\Gamma(g) \in \mathscr{E}$, then*

$$\widetilde{f}_g(x) = \frac{g(x)}{x} + \sigma^2 \widetilde{f}_{\Gamma(g)}(x) \tag{3.36}$$

and $\widetilde{f}_g'(x) = x\widetilde{f}_{\Gamma(g)}(x)$.

Proof. The corollary is a direct result of Proposition 3.3.16. □

We notice that in equation (3.28), we write $\widetilde{f}_h$ as an integral function of h, while in (3.36), $\widetilde{f}_h$ contains the derivative of h. The two expressions concerns different aspects: the smoothness and the growth rate of the functions. In fact, the previous expression concerns working with a more smooth function while in the Corollary 3.3.18, we are interested in a function with lower growth rate of g, whose auxiliary function is easier to estimate. The following simple estimation is useful.

Corollary 3.3.19 *For any integer $l \geq -1$, we have*

$$\left|\widetilde{f}_{\frac{1}{x^l}}\right| \leq \left|\widetilde{f}_{\frac{1}{|x|^l}}\right| \leq \frac{1}{|x|^{l+1}}. \tag{3.37}$$

Proof. 3) of Proposition 3.3.5 implies directly (3.37). □

We now consider function with bounded derivatives.

Proposition 3.3.20 *If h has bounded derivative, i.e. $\|h'\| \leq c$, then $|h(x)| \leq c|x| + c_1$ where c and c_1 are some constants. Then we have following estimations:*

$$|\widetilde{f}_h(x)| \leq c + \frac{c_1}{|x|}, \quad \|\widetilde{f}_h\| \leq c + \frac{\sqrt{2\pi}c_1}{2\sigma}.$$

$$|\widetilde{f}_h'(x)| \leq \frac{2c}{|x|} + \frac{c_1}{|x|^2}, \quad \|\widetilde{f}_h'(x)\| \leq \max\big(2c + c_1, \frac{2c + c_1}{\sigma^2} + \frac{\sqrt{2\pi}c_1}{2\sigma^3}\big)$$

and

$$|\widetilde{f}_h''(x)| \leq \frac{1}{\sigma^2}\big(4c + \frac{2c_1}{|x|}\big), \quad \|\widetilde{f}_h''\| \leq \max\Big(\frac{4c + 2c_1}{\sigma^2}, \frac{1}{\sigma^2}\big(c + \frac{2c + c_1}{\sigma^2} + \frac{\sqrt{2\pi}c_1}{2\sigma} + \frac{\sqrt{2\pi}c_1}{2\sigma^3}\big)\Big).$$

Proof. 1) By Proposition 3.3.5 1) 2) and Corollary 3.3.19, we have

$$|\widetilde{f}_h(x)| \leq c|\widetilde{f}_{|x|}(x)| + |\widetilde{f}_{c_1}(x)| \leq c + |\widetilde{f}_{c_1}(x)|.$$

From Corollary 3.3.10, we know that $|\widetilde{f}_{c_1}(x)| = c_1|\widetilde{f}_1(x)| \leq \frac{\sqrt{2\pi}c_1}{2\sigma}$.

2) By Corollary 3.3.18, $|\widetilde{f}_h'(x)| \leq |x||\widetilde{f}_{\Gamma(h)}(x)|$. Since

$$|\Gamma(h)| = \Big|\frac{h'(x)}{x} - \frac{h(x)}{x^2}\Big| \leq \frac{2c}{|x|} + \frac{c_1}{x^2},$$

by Proposition 3.3.5 2) and (3.37), we have

$$|\widetilde{f}_h'| \leq \frac{2c}{|x|} + \frac{c_1}{|x|^2}.$$

So when $|x| \geq 1$, $|\widetilde{f}_h'(x)| \leq 2c + c_1$. When $|x| < 1$, we use the equality $\widetilde{f}_h'(x) = \frac{1}{\sigma^2}(x\widetilde{f}_h(x) - h)$ to get

$$|\widetilde{f}_h'(x)| \leq \frac{1}{\sigma^2}\big(|\widetilde{f}_h(x)| + |h(x)|\big) \leq \frac{1}{\sigma^2}\big(2c + c_1 + \frac{\sqrt{2\pi}c_1}{2\sigma}\big)$$

3)It suffices to notice $\widetilde{f}_h''(x) = \frac{1}{\sigma^2}(\widetilde{f}_h(x) + x\widetilde{f}_h'(x) - h'(x))$ and combine 1) and 2) to complete the proof. □

We consider the function $\overline{C}_k = C_k - \bar{c}$ where C_k is the call function $C_k = (x-k)^+$ and $\bar{c} = \Phi_\sigma(C_k)$. Clearly $\overline{C}_k$ satisfies the conditions of Proposition 3.3.20 with $\|\overline{C}_k'\| = 1$. So we have $c = 1$ and $c_1 = |\overline{C}_k(0)| = |(-k)^+ - \bar{c}|$. In addition, we give in the following the estimation for $|xf_{C_k}''|$.

Corollary 3.3.21 *Let $C_k = (x-k)^+$, then*

$$\|f_{C_k}\| \leq 1 + \frac{\sqrt{2\pi}c_1}{2\sigma},$$

$$\|f_{C_k}'\| \leq \max\Big(2 + c_1, \frac{2+c_1}{\sigma^2} + \frac{\sqrt{2\pi}}{2\sigma^3}c_1\Big)$$

and

$$\|xf_{C_k}''\| \leq \max\Big[\frac{1}{\sigma^2}\Big(1 + \frac{2+c_1}{\sigma^2} + \frac{\sqrt{2\pi}c_1}{2\sigma}(1 + \frac{1}{\sigma^2})\Big), |k+\bar{c}| + 3|k| + \frac{|k|}{\sigma^2}\Big].$$

Proof. We need only to prove the last inequality. Since $\widetilde{f}_h' = x\widetilde{f}_{\Gamma(h)}$, we have $xf_{C_k}'' = x\widetilde{f}_{\Gamma(\overline{C}_k)} + x^2\widetilde{f}_{\Gamma(\overline{C}_k)}'$ where

$$\Gamma(\overline{C}_k) = \Big(\frac{(x-k)^+ - \bar{c}}{x}\Big)' = \frac{\mathbb{1}_{\{x\geq k\}}k + \bar{c}}{x^2}.$$

For the first term,

$$|\widetilde{f}_{\Gamma(\overline{C}_k)}| \leq |\widetilde{f}_{\frac{|k+\bar{c}|}{x^2}}| \leq \frac{|k+\bar{c}|}{|x|^3},$$

so when $|x| \geq 1$, $|x\widetilde{f}_{\Gamma(\overline{C}_k)}| \leq k + \bar{c}$. For the second term, we have by integration by part

$$x^2\widetilde{f}_{\Gamma(\overline{C}_k)}' = x^3\Big(\widetilde{f}_{-\mathbb{1}_{\{x\geq k\}}\frac{3k}{x^4}} + \frac{\phi_\sigma(k)}{k^2\sigma^2\phi_\sigma(x)}\mathbb{1}_{\{k\geq x>0\}} - \frac{\phi_\sigma(k)}{k^2\sigma^2\phi_\sigma(x)}\mathbb{1}_{\{k\leq x<0\}}\Big)$$

where By similar arguments as in Proposition 3.3.20, we get $|x^2\widetilde{f}_{\Gamma(\overline{C}_k)}'| \leq \frac{3|k|}{|x|^2} + \frac{k}{\sigma^2}$, which implies $|x^2f_{\Gamma(\overline{C}_k)}'| \leq 3|k| + \frac{|k|}{\sigma^2}$ when $|x| \geq 1$. When $|x| < 1$, it suffices to notice $|xf_{C_k}''| \leq \|f_{C_k}''\|$. □

3.3.4 Some remarks

Our objective is to estimate the derivatives of the function f_h and the products of the form $x^m f_h^{(l)}$. Of the two methods, the first one consists of deriving directly the expectation form $f_h(x) = \frac{\sqrt{2\pi}}{\sigma}\mathbb{E}\big[\bar{h}(Z+x)e^{-\frac{Zx}{\sigma^2}}\mathbb{1}_{\{Z>0\}}\big]$ and estimating the expectation of functions of the derivatives of h. The difficulty is that, since $\bar{h}(Z+x)e^{-\frac{Zx}{\sigma^2}}$ is of form of a product, taking derivatives increases each time the terms to estimate and it soon becomes cumbersome for higher order estimations. The second method, as we have mentioned previously, proposes to treat the problem by reducing the growth order of the function h. This method shall be discussed in a more systematic way in Chapter 4.

In this subsection, we present some other properties related to the function f_h. In the literature, the discussions mainly concentrate on the case where $\sigma = 1$. However, these results can be extended to the general case without much difficulty. The following property shows the relationship between the particular case and the general case.

Lemma 3.3.22 *For $\sigma > 0$, let $h_\sigma = \frac{1}{\sigma}(h \circ \sigma)$, then*

$$f_{h,\sigma}(x) = f_{h_\sigma,1}\Big(\frac{x}{\sigma}\Big). \tag{3.38}$$

Proof. Notice first that for any function g such that $\Phi_\sigma(g)$ exists, the following equality holds

$$\Phi_\sigma(g) = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} e^{-\frac{x^2}{2\sigma^2}}g(x)dx = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{y^2}{2}}g(\sigma y)dy = \Phi_1\big(g(\sigma y)\big) = \Phi_1\big(g \circ \sigma\big). \tag{3.39}$$

Let $x = \sigma y$, equation (3.22) implies $\sigma y f_{h,\sigma}(\sigma y) - \sigma^2 f'_{h,\sigma}(\sigma y) = h(\sigma y) - \Phi_\sigma(h)$, which follows

$$y(f_{h,\sigma}\circ\sigma)(y) - \sigma(f'_{h,\sigma}\circ\sigma)(y) = \frac{1}{\sigma}(h\circ\sigma)(y) - \Phi_1\Big(\frac{1}{\sigma}(h\circ\sigma)\Big).$$

In addition, notice that $(f_{h,\sigma}\circ\sigma)' = \sigma(f'_{h,\sigma}\circ\sigma)$, then we can rewrite the above equation as

$$yg(y) - g'(y) = h_\sigma(y) - \Phi_1(h_\sigma)$$

and we know that its solution is $g = f_{h,\sigma}\circ\sigma$. Therefore, we have $f_{h_\sigma,1} = f_{h,\sigma}\circ\sigma$. $\square$

The above result enables us to obtain some estimations directly. For example, Stein [77] has proved that $\|f''_{h,1}\| \leq 2\|h'\|$ if h is absolutely continuous. We will extend this result to $f_{h,\sigma}$ by using Lemma 3.3.22.

Proposition 3.3.23 *For any absolutely continuous function h, the solution $f_{h,\sigma}$ satisfies*

$$\|f''_{h,\sigma}\| \leq \frac{2}{\sigma^2}\|h'\|. \tag{3.40}$$

Proof. Since $f_{h,\sigma} = f_{h_\sigma,1} \circ (\sigma^{-1})$, we have $f'_{h,\sigma} = \sigma^{-1}\big(f'_{h_\sigma,1} \circ \sigma^{-1}\big)$ and $f''_{h,\sigma} = \sigma^{-2}\big(f'_{h_\sigma,1} \circ \sigma^{-1}\big)$. Then

$$\|f''_{h,\sigma}\| = \frac{1}{\sigma^2}\|f''_{h_\sigma,1} \circ \sigma^{-1}\| = \frac{1}{\sigma^2}\|f''_{h_\sigma,1}\| \leq \frac{2}{\sigma^2}\|h'_\sigma\| = \frac{2}{\sigma^2}\|h'\|.$$

□

In the normal approximation, we need to calculate the expectations of functions for normal random variables. In our case, we encounter functions of the form $x^m f_h^{(l)}$ which are not always simple and explicit to calculate. Thanks to the invariance property of normal distribution under the zero bias transformation, we here present a result which will facilitate the calculation by writing the expectation of functions containing derivatives of f_h as some polynomial functions containing h.

Proposition 3.3.24 *Let m, l be positive integers. If the l^{th}-order derivative of f_h exists, then*

$$\Phi_\sigma\big(x^m f_h^{(l)}(x)\big) = \Phi_\sigma\big(P_{m,l}(x)h(x) + Q_{m,l}(x)\big)$$

where $P_{m,l}$ and $Q_{m,l}$ are polynomial functions. When $l = 0$,

$$P_{m,0}(x) = \frac{1}{\sigma^2(m+1)}x^{m+1}; \quad Q_{m,0}(x) = \begin{cases} -\frac{\Phi_\sigma(h)}{\sigma^2(m+1)}x^{m+1}, & \textit{when } m \textit{ is impair,} \\ 0, & \textit{when } m \textit{ is pair.} \end{cases} \tag{3.41}$$

For any $l \geq 1$,

$$P_{m,l} = \frac{1}{\sigma^2}P_{m+1,l-1} - mP_{m-1,l-1}, \quad Q_{m,l} = \frac{1}{\sigma^2}Q_{m+1,l-1} - mQ_{m-1,l-1}, \ (m > 0) \tag{3.42}$$

and

$$P_{0,l} = \frac{1}{\sigma^2}P_{1,l-1} \textit{ and } Q_{0,l} = \frac{1}{\sigma^2}Q_{1,l-1}. \tag{3.43}$$

Proof. First, when $l = 0$, we have from (3.24),

$$\Phi_\sigma(x^n f_h) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}\Big(\frac{x^n}{\sigma^2}\int_{-\infty}^{x}(\Phi_\sigma(h) - h(t))e^{-\frac{t^2}{2\sigma^2}}dt\Big)dx.$$

Since for any polynomial function $P(x)$, we have

$$\lim_{x\to\pm\infty} P(x)\int_{-\infty}^{x}(\Phi_\sigma(h) - h(t))e^{-\frac{t^2}{2\sigma^2}}dt = 0.$$

Then by integration by part

$$\Phi_\sigma(x^m f_h(x)) = \Phi_\sigma\Big(\frac{x^{m+1}}{\sigma^2(m+1)}\big(h(x) - \Phi_\sigma(h)\big)\Big),$$

which implies (3.41). In particular, when m is pair, $Q_{m,0} = 0$.

When $l > 0$, we proceed by induction. We write $x^m f_h^{(l)} = (x^m f_h^{(l-1)})' - m x^{m-1} f_h^{(l-1)}$. Moreover, note that for any derivable function g, we have

$$\Phi_\sigma(g'(x)) = \frac{1}{\sigma^2} \Phi_\sigma\big(xg(x)\big). \tag{3.44}$$

Then

$$\begin{aligned}\Phi_\sigma(x^m f_h^{(l)}) &= \Phi_\sigma\big((x^m f_h^{(l-1)})'\big) - \Phi_\sigma(m x^{m-1} f_h^{(l-1)}) \\ &= \frac{1}{\sigma^2} \Phi_\sigma(x^{m+1} f_h^{(l-1)}) - m\Phi_\sigma(x^{m-1} f_h^{(l-1)})\end{aligned}$$

and we obtain (3.42). When $m = 0$, equation (3.44) implies directly the result. □

3.4 Normal approximation for conditional losses

In this section, we present our main result. We begin by some first order estimations. This has been discussed by many authors such as Stein [77], Chen [17] and Goldstein and Reinert [39]. We first revisit some of these results in our context of zero bias transformation for some regular functions and then for the indicator function. In the second subsection, we give a correction term for the normal approximation and we estimate the approximation error. In the binomial case, the error bound is of order $O(\frac{1}{n})$ after the correction. We then discuss the call function which demands more effort to prove since it does not possess second order derivative. Some numerical tests are presented to show the correction results. At last, we introduce and compare the saddle point method.

We recall the notation which shall be used in this section. Let X_i $(1 \leq i \leq n)$ be independent zero-mean random variables with variance $\sigma_i^2 > 0$ and $W = X_1 + \cdots + X_n$ with finite variance σ_W^2. We know that $W^* = W^{(I)} + X_I^*$ has the zero biased distribution where I is a random index taking values in $\{1, \cdots, n\}$ and X_i^* is independent of all $X_1, \cdots, X_n$. We denote by $(\overrightarrow{X}, \overrightarrow{X}^*) = (X_1, \cdots, X_n, X_1^*, \cdots, X_n^*)$, by $\widetilde{X}_i$ an independent duplicate of X_i and let $X_i^s = X_i - \widetilde{X}_i$. We also denote by $f_h = f_{h,\sigma_W}$.

3.4.1 Some first-ordered estimations

With the equality (3.23), the error estimation of the normal approximation can be obtained by direct Taylor expansion. However, the estimation is related to the regularity of the function f_h. In the following, we give the first ordered estimation for functions with different properties.

3.4.1.1 The regular functions

We now give first-ordered estimation for derivable functions. Proposition 3.2.6 enables us to provide a sharper bound than in Goldstein and Reinert [39].

Lemma 3.4.1 *If h has bounded derivative, then*

$$\left|\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h)\right| \leq \frac{\|h'\|}{2\sigma_W^2} \sum_{i=1}^n \mathbb{E}\left[|X_i^s|^3\right]. \tag{3.45}$$

Proof. We have by direct Taylor expansion

$$\left|\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h)\right| = \sigma_W^2 \mathbb{E}\left[|f_h'(W^*) - f_h'(W)|\right] \leq \sigma_W^2 \left\|f_h''\right\| \mathbb{E}\left[|W^* - W|\right].$$

Recall that $\left\|f_h''\right\|$ and $\mathbb{E}\left[|W^* - W|\right]$ have been estimated previously and we have $\|f_h''\| \leq \frac{2\|h'\|}{\sigma_W^2}$ (cf. Proposition 3.3.23) and $\mathbb{E}\left[|W^* - W|\right] = \frac{1}{4\sigma_W^2} \sum_{i=1}^n \mathbb{E}\left[|X_i^s|^3\right]$, which follows (3.45). □

The condition in the previous lemma can be relaxed. Instead of the boundedness condition of h', we now suppose that h' increases linearly.

Lemma 3.4.2 *If the derivative of h is of linear increasing order, i.e., $|h'(x)| \leq a|x| + b$, then*

$$\begin{aligned} |\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h)| \leq & \frac{b_1}{4} \sum_{i=1}^n \mathbb{E}\left[|X_i^s|^3\right] + a_1 \sigma_W^2 \sqrt{\frac{1}{6} \sum_{i=1}^n \mathbb{E}[|X_i^s|^4]} \quad \sim O(\frac{1}{\sqrt{n}}) \\ & + \frac{a_1}{12} \sum_{i=1}^n \mathbb{E}[|X_i^s|^4] \qquad \sim O(\frac{1}{n}) \end{aligned}$$

where a_1 and b_1 are some constants.

Proof. Since h' is of linear increasing order, f_h'' is also of linear increasing order and there exist some constants a_1 and b_1 such that $|f_h''(x)| \leq a_1|x| + b_1$. Hence

$$\begin{aligned} \mathbb{E}[|h(W) - \Phi_{\sigma_W}(h)|] &\leq \sigma_W^2 \mathbb{E}[|f_h''(W + \xi(X_I^* - X_I))(X_I^* - X_I)|] \\ &\leq \sigma_W^2 \mathbb{E}\left[|a_1(W + \xi(X_I^* - X_I)) + b_1||X_I^* - X_I|\right] \\ &\leq \sigma_W^2 \Big(a_1 \mathbb{E}[|W||X_I^* - X_I|] + \frac{a_1}{2} \mathbb{E}[(X_I^* - X_I)^2] + b_1 \mathbb{E}[|X_I^* - X_I|]\Big). \end{aligned}$$

We estimate the first term by the Cauchy-Schwarz inequality,

$$\mathbb{E}[|W||X_I^* - X_I|] \leq \sqrt{\mathbb{E}[W^2]} \sqrt{\mathbb{E}[|X_I^* - X_I|^2]} = \sqrt{\frac{1}{6} \sum_{i=1}^n \mathbb{E}[|X_i^s|^4]}.$$

The other terms are easily estimated by Proposition 3.2.6. □

Remark 3.4.3 1. Lemma 3.4.2 requires the existance of the fourth order moment of X_i. This is due to the linear increasing property of the derivative of h.

2. Lemma 3.4.2 can be extended to the case where h' is of polynomial increasing order, that is, if there exist $x_0 > 0$ and some constant c such at $h'(x) \leq c|x|^n$ for $|x| > x_0$. However, it's necessary that higher moments of X_i exist.

3.4.1.2 The indicator function

The approximation error of the indicator function $\mathbb{1}_{\{x \leq k\}}$ is estimated by the Berry-Esseen inequality. We here introduce a method based on the Stein's method and the zero bias transformation to obtain the estimation. The key tool is a concentration inequality of Chen and Shao [17], which is also essential for the estimation of the call function whose derivative is the indicator function. We give a proof of the concentration inequality by writing the zero bias transformation, which is coherent in our context. Our objective here is not to find the optimal estimation constant.

To prove the concentration inequality, the idea is to majorize $\mathbb{P}(a \leq W \leq b)$ by $\mathbb{P}(a - \varepsilon \leq W^* \leq b + \varepsilon)$ up to a small error with a suitable ε.

Proposition 3.4.4 *(Chen and Shao) For any real a and b, we have*

$$\mathbb{P}(a \leq W \leq b) \leq \frac{b-a}{\sigma_W} + \frac{\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]}{\sigma_W^3} + \frac{\Big(\sum_{i=1}^n \frac{\sigma_i}{\sqrt{2}}\mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}}{2\sigma_W^2}.$$

Proof. Let f' be the indicator function $f'(x) = \mathbb{1}_{[a-\varepsilon, b+\epsilon]}(x)$ where ε is a positive constant. One primitive function is given by $f(x) = \int_{(a+b)/2}^x f'(t)dt$, which is bounded by $|f(x)| \leq \varepsilon + \frac{b-a}{2}$. Using the zero bias transformation, we have

$$\sigma_W^2 \mathbb{E}[\mathbb{1}_{[a-\varepsilon, b+\varepsilon]}(W^*)] = \mathbb{E}[Wf(W)] \leq \sigma_W\big(\varepsilon + \frac{b-a}{2}\big).$$

On the other hand,

$$\begin{aligned}\mathbb{P}(a - \varepsilon \leq W^* \leq b + \varepsilon) &\geq \mathbb{P}(a \leq W \leq b, |X_I - X_I^*| \leq \varepsilon)\\ &= \mathbb{P}(a \leq W \leq b)\mathbb{P}(|X_I^* - X_I| \leq \varepsilon) + \mathrm{cov}(\mathbb{1}_{\{a \leq W \leq b\}}, \mathbb{1}_{\{|X_I^* - X_I| \leq \varepsilon\}}).\end{aligned}$$

As shown by (3.20),

$$\mathrm{cov}(\mathbb{1}_{\{a \leq W \leq b\}}, \mathbb{1}_{\{|X_I^* - X_I| \leq \varepsilon\}}) \geq -\frac{1}{2\sigma_W^2}\big(\sum_{i=1}^n \frac{\sigma_i}{4\sqrt{2}}\mathbb{E}[|X_i^s|^3]\big)^{\frac{1}{2}}$$

Therefore, we get the following inequality

$$\varepsilon + \frac{b-a}{2} \geq \sigma_W \mathbb{P}(a \leq W \leq b)\mathbb{P}(|X_I^* - X_I| \leq \varepsilon) - \frac{1}{2\sigma_W}\Big(\sum_{i=1}^n \frac{\sigma_i}{4\sqrt{2}}\mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}. \quad (3.46)$$

Denote by $A_\varepsilon = \sigma_W \mathbb{P}(|X_I^* - X_I| \leq \varepsilon)$,

$$B = \frac{1}{2\sigma_W}\big(\sum_{i=1}^{n}\frac{\sigma_i}{4\sqrt{2}}\mathbb{E}[|X_i^s|^3]\big)^{\frac{1}{2}}$$

and $C_\varepsilon = \varepsilon + \frac{b-a}{2}$, we can rewrite the above inequality as $\mathbb{P}(a \leq W \leq b)A_\varepsilon \leq B + C_\varepsilon$ and we are interested in majorizing $\frac{B+C_\varepsilon}{A_\varepsilon}$.

In fact, by Corollary 3.2.13,

$$A_\varepsilon \geq \sigma_W - \frac{1}{4\sigma_W \varepsilon}\sum_{i=1}^{n}\mathbb{E}[|X_i^s|^3].$$

So we shall choose ε such that C_ε is of the same order of B. Let $\sum_{i=1}^n \frac{\mathbb{E}[|X_i^s|^3]}{4\varepsilon} = \frac{1}{2}\sigma_W^2$. That is,

$$\varepsilon = \frac{1}{2\sigma_W^2}\sum_{i=1}^{n}\mathbb{E}[|X_i^s|^3].$$

So $A_\varepsilon \geq \frac{\sigma_W}{2}$ and

$$C_\varepsilon = \frac{b-a}{2} + \frac{\sum_{i=1}^n \mathbb{E}[|X_i^s|^3]}{2\sigma_W^2},$$

which follows

$$\mathbb{P}(a \leq W \leq b) \leq \frac{b-a}{\sigma_W} + \frac{\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]}{\sigma_W^3} + \frac{\Big(\sum_{i=1}^n \frac{\sigma_i}{\sqrt{2}}\mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}}{2\sigma_W^2}.$$

□

In the following, we shall use sometimes the upper bound of $\mathbb{P}(a \leq W^{(i)} \leq b)$. Since $W^{(i)}$ also the sum of several random variables, the above proposition applies of course directly to $W^{(i)}$ by removing the variate i in the sum terms of the right-hand side. However, for the simplicity of writing, we prefer keep all the summand terms. To this end, we shall use the independence property between $W^{(i)}$ and X_i, X_i^* to get another concentration inequality. Here again, our objective is not the optimal estimation.

Corollary 3.4.5 *For any real a and b, we have*

$$\mathbb{P}(a \leq W^{(i)} \leq b) \leq \frac{2}{\sigma_W}\big((b-a) + 4\sigma_i\big) + \frac{2\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]}{\sigma_W^3} + \frac{\Big(\sum_{i=1}^n \frac{\sigma_i}{\sqrt{2}}\mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}}{\sigma_W^2}.$$

Proof. For any $\varepsilon > 0$, we have $\mathbb{P}(a \leq W^{(i)} \leq b, |X_i| \leq \varepsilon) \leq \mathbb{P}(a - \varepsilon \leq W \leq b + \varepsilon)$, then by Markov's inequality and the independence between $W^{(i)}$ and X_i,

$$\mathbb{P}(a \leq W^{(i)} \leq b) \leq \frac{\mathbb{P}(a-\varepsilon \leq W \leq b+\varepsilon)}{1 - \frac{\mathbb{E}[|X_i|]}{\varepsilon}}.$$

We choose $\varepsilon = 2\mathbb{E}[|X_i|]$ and apply Proposition 3.4.4 to end the proof. $\square$

Now we give the approximation estimation of the indicator function. The difficulty lies in the irregularity of the function. We shall use, on one hand, the nearness between W and W^* and on the other hand, the fact that the zero bias transformation enables us to work with a more regular function.

Proposition 3.4.6 *Let $I_\alpha = \mathbb{1}_{\{x\leq\alpha\}}$, then*

$$\begin{aligned}|\mathbb{E}[I_\alpha(W)] - \mathcal{N}_{\sigma_W}(\alpha)| \leq \frac{c}{4\sigma_W^2}\sum_{i=1}^n \mathbb{E}[|X_i^s|^3] + \sum_{i=1}^n \frac{2\sigma_i^3}{\sigma_W^3}\Big(\frac{\mathbb{E}[|X_i^s|^3]}{4\sigma_i^3}+4\Big) \\ + \frac{2\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]}{\sigma_W^3} + \frac{\Big(\sum_{i=1}^n \frac{\sigma_i}{\sqrt{2}}\mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}}{\sigma_W^2}\end{aligned} \tag{3.47}$$

where $c = \|f_{I_\alpha}\| + \|xf'_{I_\alpha}\|$.

Proof. We write $I_\alpha(W) - \mathcal{N}_{\sigma_W}(\alpha)$ as the sum of two difference terms, i.e.

$$I_\alpha(W) - \mathcal{N}_{\sigma_W}(\alpha) = \big(I_\alpha(W) - I_\alpha(W^*)\big) + \big(I_\alpha(W^*) - \Phi_{\sigma_W}(\alpha)\big).$$

We shall estimate the two terms respectively. For the first term, since

$$\mathbb{1}_{\{x+y\leq\alpha\}} - \mathbb{1}_{\{x+z\leq\alpha\}} = \mathbb{1}_{\{\alpha-\max(y,z)<x\leq\alpha-\min(y,z)\}}, \tag{3.48}$$

then

$$\mathbb{E}[I_\alpha(W^{(i)}+X_i) - I_\alpha(W^{(i)}+X_i^*)] = \mathbb{P}(\alpha - \max(X_i,X_i^*) < W^{(i)} \leq \alpha - \min(X_i,X_i^*)).$$

Since $W^{(i)}$ and X_i, X_i^* are independent, using Corollary 3.4.5, we obtain

$$\begin{aligned}\mathbb{E}\big[|I_\alpha(W) - I_\alpha(W^*)|\big] \leq \sum_{i=1}^n \frac{2\sigma_i^3}{\sigma_W^3}\Big(\frac{\mathbb{E}[|X_i^s|^3]}{4\sigma_i^3}+4\Big) \\ + \sum_{j=1}^n \frac{\sigma_j^2}{\sigma_W^2}\Big(\frac{2\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]}{\sigma_W^3} + \frac{\Big(\sum_{i=1}^n \frac{\sigma_i}{\sqrt{2}}\mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}}{\sigma_W^2}\Big).\end{aligned}$$

For the second term, by the zero bias transformation

$$\mathbb{E}[I_\alpha(W^*)] = -\frac{1}{\sigma_W^2}\mathbb{E}[W(\alpha - W)^+].$$

Denote the primitive of I_α by $G_I(x) = -(\alpha - x)^+$ and by $\widetilde{G}_I(x) = xG_I(x)$. Then we shall estimate

$$\frac{1}{\sigma_W^2}\big(\mathbb{E}[\widetilde{G}_I(W)] - \Phi_{\sigma_W}(\widetilde{G}_I)\big) = \mathbb{E}[f'_{\widetilde{G}_I}(W^*) - f'_{\widetilde{G}_I}(W)].$$

Notice that $I_\alpha = \Gamma(\widetilde{G}_I)$, then by Corollary 3.3.18, we have $f'_{\widetilde{G}_I} = xf_{I_\alpha}$. Hence $|f''_{\widetilde{G}_I}(x)| \leq |f_{I_\alpha}| + |xf'_{I_\alpha}| \leq c$, where c is estimated in Corollary 3.3.13 and $c \leq \frac{\sqrt{2\pi}}{2\sigma}\left(\frac{1}{e}+1\right) + \frac{|\alpha|}{\sigma^2}$. Then

$$\mathbb{E}[f'_{\widetilde{G}_I}(W^*) - f'_{\widetilde{G}_I}(W)] \leq c\mathbb{E}[|W^* - W|] = \frac{c}{4\sigma_W^2}\sum_{i=1}^n \mathbb{E}[|X_i^s|^3].$$

We complete the proof by combining the estimations of the two terms. □

Combining the above results for the regular function case and the indicator function case, we can estimate the approximation error for a larger class of functions of finite variation under some conditions.

Proposition 3.4.7 *If the function h is of local finite variation with the derivative of the continuous part h_1 of linear increasing order and the pure jump part having finite total jumps, then*

$$\begin{aligned}|\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h)| \leq & \sum_{t\in\mathbb{R}} |\Delta h(t)| \sum_{j=1}^{+\infty} B(W, t_j) \\ & + \frac{b_1}{4}\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big] + a_1\sigma_W^2\sqrt{\frac{1}{6}\sum_{i=1}^n \mathbb{E}[|X_i^s|^4]} + \frac{a_1}{12}\sum_{i=1}^n \mathbb{E}[|X_i^s|^4]\end{aligned}$$

where $B(W, t_j)$ is the normal approximation error bound in equation (3.47) for the indicator function $1\!\!1_{\{W\leq t_j\}}$ with t_j being the jump points of the function h and a_1 and b_1 are constants such that $|f''_{h_1}(x)| \leq a_1|x| + b_1$.

Proof. We write the function h as

$$h = h_1 + \sum_{t\leq x} \Delta h(t),$$

where h_1 is absolutely continuous and $\Delta h(t) = h(t+) - h(t-)$. By the linearity of $\Phi_{\sigma_W}(h)$, we have

$$|\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h)| \leq |\mathbb{E}[h_1(W)] - \Phi_{\sigma_W}(h_1)| + \sum_{t\in\mathbb{R}} |\Delta h(t)||\mathbb{E}[1\!\!1_{\{W\leq t\}}] - \mathcal{N}_{\sigma_W}(t)|.$$

Then it suffices to apply Lemma 3.4.2 and Proposition 3.4.6. □

3.4.2 Correction for asymmetric normal approximation

We now propose a first order approximation correction to improve the approximation accuracy.

Theorem 3.4.8 *Let $X_1, \cdots, X_n$ be random variables such that $\mathbb{E}[X_i^4]$ $(i = 1, \cdots, n)$ exist. If the function h is Lipschitz and if f_h has bounded third order derivative, then the normal approximation $\Phi_{\sigma_W}(h)$ of $\mathbb{E}[h(W)]$ has corrector*

$$C_h = \frac{1}{\sigma_W^2}\mathbb{E}[X_I^*]\Phi_{\sigma_W}\left(\Big(\frac{x^2}{3\sigma_W^2} - 1\Big)xh(x)\right). \tag{3.49}$$

Recall that $\mathbb{E}[X_I^*] = \frac{1}{2\sigma_W^2}\sum_{i=1}^n \mathbb{E}[X_i^3]$. The corrected error is bounded by

$$\begin{aligned}&\Big|\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h) - C_h\Big| \\ \leq\ &\|f_h^{(3)}\|\left(\frac{1}{12}\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^4\big] + \frac{1}{4\sigma_W^2}\Big|\sum_{i=1}^n \mathbb{E}[X_i^3]\Big|\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big] + \frac{1}{\sigma_W}\sqrt{\sum_{i=1}^n \sigma_i^6}\right).\end{aligned}$$

Proof. The normal approximation error is given by equation (3.23). Then taking first order Taylor expansion, we have

$$\begin{aligned}&\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h) = \sigma_W^2\mathbb{E}[f_h'(W^*) - f_h'(W)] \\ = &\ \sigma_W^2\mathbb{E}[f_h''(W)(W^* - W)] + \sigma_W^2\mathbb{E}\Big[f_h^{(3)}\big(\xi W + (1-\xi)W^*\big)\xi(W^* - W)^2.\Big]\end{aligned} \tag{3.50}$$

where ξ is a uniform variable on $[0, 1]$ independent of all X_i and X_i^*. First, we notice that the remaining term is bounded by

$$\mathbb{E}\Big[\big|f_h^{(3)}\big(\xi W + (1-\xi)W^*\big)\xi\big|(W^* - W)^2\Big] \leq \frac{\|f_h^{(3)}\|}{2}\mathbb{E}[(W^* - W)^2].$$

Then we have by Corollary 3.2.12

$$\sigma_W^2\Big|\mathbb{E}\Big[f_h^{(3)}\big(\xi W + (1-\xi)W^*\big)\xi(W^* - W)^2\Big]\Big| \leq \frac{\|f_h^{(3)}\|}{12}\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^4\big]. \tag{3.51}$$

Second, we consider the first term of equation (3.50). Since X_I^* is independent of W, we have

$$\mathbb{E}[f_h''(W)(W^* - W)] = \mathbb{E}[f_h''(W)(X_I^* - X_I)] = \mathbb{E}[X_I^*]\mathbb{E}[f_h''(W)] - \mathbb{E}[f_h''(W)X_I]. \tag{3.52}$$

For the second term $\mathbb{E}[f_h''(W)X_I]$ of (3.52), we have by Proposition 3.2.16

$$\Big|\mathbb{E}[f_h''(W)X_I]\Big| \leq \frac{1}{\sigma_W^2}\sqrt{\mathrm{Var}[f_h''(W)]}\sqrt{\sum_{i=1}^n \sigma_i^6}. \tag{3.53}$$

Notice that $\mathrm{Var}[f_h''(W)] = \mathrm{Var}[f_h''(W) - f_h''(0)] \leq \mathbb{E}[(f_h''(W) - f_h''(0))^2] \leq \|f_h^{(3)}\|^2\sigma_W^2$. Therefore

$$\left|\mathbb{E}[f_h''(W)X_I]\right| \leq \frac{\|f_h^{(3)}\|}{\sigma_W}\sqrt{\sum_{i=1}^n \sigma_i^6}$$

For the first term $\mathbb{E}[X_I^*]\mathbb{E}[f_h''(W)]$ of (3.52), we write it as the sum of two parts

$$\mathbb{E}[X_I^*]\mathbb{E}[f_h''(W)] = \mathbb{E}[X_I^*]\Phi_{\sigma_W}(f_h'') + \mathbb{E}[X_I^*]\mathbb{E}[f_h''(W) - \Phi_{\sigma_W}(f_h'')].$$

We apply Lemma 3.4.1 to the second part and get

$$\left|\mathbb{E}[X_I^*]\Big(\mathbb{E}[f_h''(W)] - \Phi_{\sigma_W}(f_h'')\Big)\right| \leq \frac{\|f_h^{(3)}\|}{4\sigma_W^4}\left|\sum_{i=1}^n \mathbb{E}[X_i^3]\right|\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]. \tag{3.54}$$

Then, it suffices to write

$$\begin{aligned}\mathbb{E}[h(w)] - \Phi_{\sigma_W}(h) = \sigma_W^2\Big(&\mathbb{E}[X_I^*]\Phi_{\sigma_W}(f_h'') + \mathbb{E}[X_I^*]\big[\mathbb{E}[f_h''(W)] - \Phi_{\sigma_W}(f_h'')\big] - \mathbb{E}[f_h''(W)X_I]\Big)\\ &+ \sigma_W^2\mathbb{E}\Big[f_h^{(3)}(\xi W + (1-\xi)W^*)\xi(W^* - W)^2\Big].\end{aligned} \tag{3.55}$$

Combining (3.51), (3.53) and (3.54), we deduce the error bound. Finally, we apply Proposition 3.3.24 to obtain

$$C_h = \sigma_W^2\mathbb{E}[X_I^*]\Phi_{\sigma_W}(f_h'') = \frac{1}{\sigma_W^2}\mathbb{E}[X_I^*]\Phi_{\sigma_W}\Big(\big(\frac{x^2}{3\sigma_W^2} - 1\big)xh(x)\Big).$$

□

Remark 3.4.9 1. We notice that C_h contains two parts. On one hand, $\mathbb{E}[X_I^*]$ depends only on the variables $X_1, \cdots, X_n$. On the other hand, the term containing Φ_{σ_W} depends only on the function h itself. Hence we can study the two parts separately. Moreover, it is worth noting that both terms are easy to calculate.

2. For the binomial case, the corrected approximation error bound is of order $O(\frac{1}{n})$. If, in addition, $\mathbb{E}[X_i^3] = 0$ for any $i = 1, \cdots, n$, then the error of the approximation without correction is automatically of order $O(\frac{1}{n})$. This result has been mentioned in Feller [32] concerning the Edgeworth expansion and has been discussed in Goldstein and Reinert [39].

3. In the symmetric case, $\mathbb{E}[X_I^*] = 0$, then $C_h = 0$ for any function h. Therefore, the corrector C_h is most effective for the asymmetric case in the sense that after correction, the asymmetric approximations obtain the same order of the approximation error as in the symmetric case.

In some cases, it is difficult to calculate the explicit form of the function f_h'' and the expectation $\Phi_\sigma(f_h'')$. However, it is possible to simplify the calculation with different forms of $\Phi_{\sigma_W}(f_h'')$ by using the normal function property (3.7) and by the relationship between the functions h and f_h implied by the Stein's equation.

Corollary 3.4.10 *Under the condition of Theorem 3.4.8, we have following equivalent forms of C_h:*

1. $C_h = \sigma_W^2 \mathbb{E}[X_I^*]\Phi_{\sigma_W}(f_h'')$;
2. $C_h = \mathbb{E}[X_I^*]\Phi_{\sigma_W}(xf_h'(x))$;
3. $C_h = \frac{\mathbb{E}[X_I^*]}{3}\left[\frac{1}{\sigma_W^2}\Phi_{\sigma_W}(x^2h'(x)) - \Phi_{\sigma_W}(h')\right]$;
4. $C_h = \frac{\mathbb{E}[X_I^*]}{3}\Phi_{\sigma_W}(xh''(x))$.

Proof. 1) is obtained in the proof of Theorem 3.4.8.
2) is direct by 1) using $\Phi_\sigma(g') = \frac{1}{\sigma^2}\Phi_\sigma(xg)$ and similarly, 3) is direct by (3.49) and 4) is by 3). □

Note that h'' exists and is bounded since f_h has bounded third order derivative. Hence, we can calculate C_h according to the explicit form of the function h and f_h to simplify the computation.

Remark 3.4.11 The formula 4) of the above corollary shows that if h'' is an even function, then $\Phi_{\sigma_W}(xh'') = 0$ and the corrector C_h vanishes. In particular, for the polynomial functions of even order $h(x) = x^{2l}$ where l is a positive integer, $C_h = 0$.

3.4.2.1 Some examples

We now consider $W = \sum_{i=1}^n X_i$ where X_i are independent but non-identical Bernoulli random variables which follow $\mathcal{B}_{\gamma_i}(q, -p)$. Denote by σ_i^2 the variance of X_i, then $\gamma_i = \frac{\sigma_i}{\sqrt{p(1-p)}}$. Here W is a zero-mean random variable with finite variance σ_W^2 and

$$\mathbb{E}[X_I^*] = \frac{1}{2\sigma_W^2}\sum_{i=1}^n \mathbb{E}[X_i^3] = \frac{1}{2\sigma_W^2}\sum_{i=1}^n \sigma_i^3 \frac{1-2p}{\sqrt{p(1-p)}}.$$

In particular, if $X_1, \cdots, X_n$ follow identical distribution such that $\sigma_i = \frac{\sigma}{\sqrt{n}}$, then W is a zero-mean binomial random variable and $\mathbb{E}[X_I^*] = \mathbb{E}[X_1^*] = \frac{\sigma_W(1-2p)}{2\sqrt{np(1-p)}}$, moreover,

$$C_h = \frac{1}{2\sigma_W}\frac{1-2p}{\sqrt{np(1-p)}}\Phi_{\sigma_W}\left(\left(\frac{x^2}{3\sigma_W^2} - 1\right)xh(x)\right).$$

When $p = \frac{1}{2}$, $C_h = 0$. In addition, for a given p, the corrector is of order $\frac{1}{\sqrt{n}}$.

We give some simple examples of function h.

Example 3.4.12 $h(x) = x^2$. The corrector $C_h = 0$ by Corollary 3.4.10. In fact, it is easy to see that $\mathbb{E}[W^2] = \Phi_{\sigma_W}(h) = \sigma_W^2$, so there is no need of correction.

Example 3.4.13 $h(x) = x^3$. The correction is given by

$$C_h = \frac{\mathbb{E}[X_I^*]}{3}\left[\frac{1}{\sigma_W^2}\Phi_{\sigma_W}(3x^4) - \Phi_{\sigma_W}(3x^2)\right] = \sum_{i=1}^n \mathbb{E}[X_i^3]. \tag{3.56}$$

On the other hand, $\mathbb{E}[W^3] - \Phi(h) = \mathbb{E}\big[(\sum_{i=1}^n X_i)^3\big]$. Since $X_1, \cdots, X_n$ are independent, the corrected approximation is exact.

3.4.3 "Call" function

In this subsection, we concentrate on the "call" function $C_k(x) = (x-k)^+$. This is a Lipschitz function with $C_k'(x) = \mathbb{1}_{\{x>k\}}$. Notice that C_k'' exists only in distribution sense. So we can no longer majorize the error of the the corrected approximation via the norm $\|f_h^{(3)}\|$. However, we calculate

$$\frac{1}{3\sigma_W^2}\Phi_\sigma\Big(\mathbb{1}_{\{x\geq k\}}\big(\frac{x^2}{\sigma_W^2} - 1\big)\Big) = \frac{1}{3\sigma_W^4}\int_k^\infty x^2\phi_\sigma(x)dx - (1 - \mathcal{N}_\sigma(k)) = \frac{k}{3}\phi_\sigma(k),$$

and the corrector is given by

$$C_{(x-k)^+} = \frac{1}{3}\mathbb{E}[X_I^*]k\phi_{\sigma_W}(k). \tag{3.57}$$

In particular, for the homogeneous case,

$$C_{(x-k)^+} = \frac{\sigma_W(1-2p)}{6\sqrt{np(1-p)}}k\phi_{\sigma_W}(k).$$

Remark 3.4.14 When the strike $k = 0$, the correction disappears, which means that the error bound of the normal approximation for the function $h(x) = x^+$ is automatically of order $O(\frac{1}{n})$. Heuristically, this property can be shown by Remark 3.4.11.

We now provide the error estimation of the corrected normal approximation for the call function. The following theorem shows that although the conditions of Theorem 3.4.8 are not satisfied here, (3.57) remains the approximation corrector and the approximation error is estimated. With the corrector C_h, we are interested in a error estimation of order $O(\frac{1}{n})$ in the binomial case. As we have stated above, the difficulty is mainly due to the fact that h'' and $f_h^{(3)}$ do not exist and that the proof of the Theorem 3.4.8 is no longer valid. We now present an alternative proof for the call function. In fact, we write f_h'' as some more regular functions by the Stein's equation and we shall use some previously obtained first order estimations.

Proposition 3.4.15 *Let $X_1, \cdots, X_n$ be independent zero-mean random variables such that $\mathbb{E}[X_i^4]$ $(i = 1, \cdots, n)$ exist. Then the error of the corrected normal approximation for the function $C_k(x) = (x-k)^+$ is bounded by*

$$
\begin{aligned}
&\left|\mathbb{E}[(W-k)^+] - \Phi_{\sigma_W}((x-k)^+) - C_{(x-k)^+}\right| \\
&\leq \frac{1}{\sigma_W^2}\sum_{i=1}^n \Big(\frac{\mathbb{E}[|X_i^s|^4]}{3} + \sigma_i \mathbb{E}[|X_i^s|^3]\Big) \\
&+ \frac{1}{4\sigma_W^2}\sum_{i=1}^n \mathbb{E}[|X_i^s|^3]\Big(\frac{2\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]}{\sigma_W^3} + \frac{\Big(\sum_{i=1}^n \sigma_i \mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}}{\sqrt{2}\sigma_W^2}\Big) \\
&+ \mathrm{Var}[f''_{C_k}(W)]^{\frac{1}{2}}\Big(\sum_{i=1}^n \sigma_i^6\Big)^{\frac{1}{2}} + \frac{1}{4\sigma_W^2}\sum_{i=1}^n \mathbb{E}[|X_i^s|^3]\Big(B(W,k) + \frac{c}{2\sigma_W^2}\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]\Big)
\end{aligned}
\tag{3.58}
$$

where $c = 2\|f'_{C_k}\| + \|xf''_{C_k}\|$ and $B(W,k)$ is the normal approximation error bound for the indicator function $\mathbb{1}_{\{W\leq k\}}$.

Proof. Similar with the equation (3.55), we first decompose the corrected error $\mathbb{E}[C_k(W)] - \Phi_{\sigma_W}(C_k) - \sigma_W^2\Phi_{\sigma_W}(f''_{C_k})\mathbb{E}[X_I^*]$ as the sum of three terms except that we replace the third-ordered derivative term, i.e.

$$
\begin{aligned}
&\mathbb{E}[C_k(W)] - \Phi_{\sigma_W}(C_k) - \sigma_W^2\mathbb{E}[X_I^*]\Phi_{\sigma_W}(f''_{C_k}) \\
&= \sigma_W^2\mathbb{E}[X_I^*]\Big(\mathbb{E}[f''_{C_k}(W)] - \Phi_{\sigma_W}(f''_{C_k})\Big) && (3.59)\\
&- \sigma_W^2\mathbb{E}[f''_{C_k}(W)X_I] && (3.60)\\
&+ \sigma_W^2\mathbb{E}[f'_{C_k}(W^*) - f'_{C_k}(W) - f''_{C_k}(W)(X_I^* - X_I)]. && (3.61)
\end{aligned}
$$

We then estimate each term respectively.

For (3.59), we use the Stein's equation $f'_h(x) = \frac{1}{\sigma_W^2}(xf_h(x) - \bar{h}(x))$ to get

$$f''_{C_k} = \sigma_W^{-2}(f_{C_k}(x) + xf'_{C_k}(x) - C'_k(x)).$$

Then

$$\sigma_W^2|\mathbb{E}[f''_{C_k}(W)] - \Phi_{\sigma_W}(f''_{C_k})| \leq |\mathbb{E}[g(W)] - \Phi_{\sigma_W}(g)| + |\mathbb{E}[\mathbb{1}_{\{W\leq k\}}] - \mathcal{N}_{\sigma_W}(k)|$$

where $g = f_{C_k} + xf'_{C_k}$ is a derivable function and $\|g'\| \leq 2\|f'_{C_k}\| + \|xf''_{C_k}\| = c$, where c has estimated in Corollary 3.3.15 and Corollary 3.3.21 by the two methods respectively. Therefore,

$$\left|\sigma_W^2(\mathbb{E}[f''_{C_k}(W)] - \Phi_{\sigma_W}(f''_{C_k}))\right| \leq B(W,k) + \frac{c}{2\sigma_W^2}\sum_{i=1}^n \mathbb{E}[|X_i^s|^3].$$

The second term (3.60) is bounded by $\sigma_W^2 \mathbb{E}[f''_{C_k}(W) X_I] \le \sqrt{\mathrm{Var}[f''_{C_k}(W)]}\sqrt{\sum_{i=1}^n \sigma_i^6}$ by Proposition 3.2.16. For (3.61), we use again the Stein's equation to write $f'_{C_k} = \sigma_W^{-2}(xf_{C_k} - \overline{C}_k)$. Denote by $G(x) = xf_{C_k}(x)$. Notice that G is the primitive function of g, then

$$
\begin{aligned}
&\sigma_W^2 \mathbb{E}[|f'_{C_k}(W^*) - f'_{C_k}(W) - f''_{C_k}(W)(X_I^* - X_I)|] \\
&\le \mathbb{E}[|G(W^*) - G(W) - g(W)(X_I^* - X_I)|] \\
&+ \mathbb{E}[|C_k(W^*) - C_k(W) - C_k'(W)(X_I^* - X_I)|].
\end{aligned} \tag{3.62}
$$

The first term of (3.62) is bounded by $c\mathbb{E}[(X_I^* - X_I)^2]$. For the second term of (3.62), notice that the call function satisfies

$$
|C_k(x+a) - C_k(x+b) - (a-b)C_k'(x+b)| \le \mathbb{1}_{\{k-\max(a,b) \le x \le k - \min(a,b)\}} |a-b|
$$

Hence

$$
\begin{aligned}
&\mathbb{E}[|C_k(W^*) - C_k(W) - C_k'(W)(X_I^* - X_I)|] \\
&\le \mathbb{E}[|X_I^* - X_I| \mathbb{1}_{\{k - \max(X_I^*, X_I) \le W^{(I)} \le k - \min(X_I^*, X_I)\}}].
\end{aligned}
$$

Since $W^{(i)}$ is independent of X_i and X_i^*, we have by using the concentration inequality that

$$
\begin{aligned}
&\mathbb{E}\big[|X_i^* - X_i| \mathbb{1}_{\{k-\max(X_i^*, X_i) \le W^{(i)} \le k - \min(X_i^*, X_i)\}}\big] \\
&= \mathbb{E}\big[|X_i^* - X_i| \mathbb{E}[\mathbb{1}_{\{k-\max(X_i^*, X_i) \le W^{(i)} \le k - \min(X_I^*, X_I)\}} | X_i, X_i^*]\big] \\
&\le \mathbb{E}[|X_i^* - X_i| B_1(W^i, |X_i^* - X_i|)]
\end{aligned}
$$

where $B_i(W^{(i)}, |X_i^* - X_i|)$ is the error bound for $\mathbb{P}(k - \max(X_i^*, X_i) \le W^{(i)} \le k - \min(X_i^*, X_i))$. Therefore,

$$
\begin{aligned}
&\mathbb{E}[|C_k(W^*) - C_k(W) - C_k'(W)(X_I^* - X_I)|] \\
&\le \frac{1}{\sigma_W^2} \sum_{i=1}^n \Big(\frac{\mathbb{E}[|X_i^s|^4]}{3} + \sigma_i \mathbb{E}[|X_i^s|^3] \Big) \\
&+ \frac{1}{4\sigma_W^2} \sum_{j=1}^n \mathbb{E}[|X_j^s|^3] \Big(\frac{2\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]}{\sigma_W^3} + \frac{\Big(\sum_{i=1}^n \sigma_i \mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}}{\sqrt{2}\sigma_W^2} \Big).
\end{aligned}
$$

Finally, combining the three terms, we get the estimation (3.58). □

Remark 3.4.16 1. In the homogeneous case where X_i are i.i.d asymmetric Bernoulli variables, the error bound is of order $O(\frac{1}{n})$.

2. The symmetric case when $p = \frac{1}{2}$ have been studied by several authors in the context of computing the call option prices by the binomial tree model. Diener and Diener [25] and Gobet [37] have proved that the convergence speed towards the Black-Scholes price is of order $O(\frac{1}{n})$. In addition, they pointed out and discussed the oscillatory behavior when n tends to infinity. Proposition 3.4.15 applied in the homogeneous case provides another proof for the convergence speed. However, it concerns no explanation of the oscillation.

3.4.3.1 Numerical results

We now provide numerical results of different tests. We take $X_1, \cdots, X_n$ to be independent asymmetric Bernoulli random variables. We shall compare the expectation of $\mathbb{E}[h(W)]$ and its normal approximation with and without our correction. It is shown that the corrector improves the approximation.

1. Call function: the homogeneous case, (Figure 3.1 and Figure 3.2).

   In each of the following two figures, three curves are presented which are

   (a) $\mathbb{E}[(W-k)^+]$ where $W = X_1 + \cdots + X_n$ is a binomial random variable of expectation zero and variance 1. More precisely, X_i are identically distributed asymmetric Bernoulli random variables and $X_1 \sim \mathcal{B}_\gamma(1-p, -p)$ such that $\mathrm{Var}(X_1) = \gamma^2 p(1-p) = \frac{1}{n}$. Therefore $\gamma = \frac{1}{\sqrt{np(1-p)}}$. Or in other words, let $H \sim B(n,p)$ be a standard binomial random variable of variance $np(1-p) = 1$, then $W = \gamma(H - np)$;

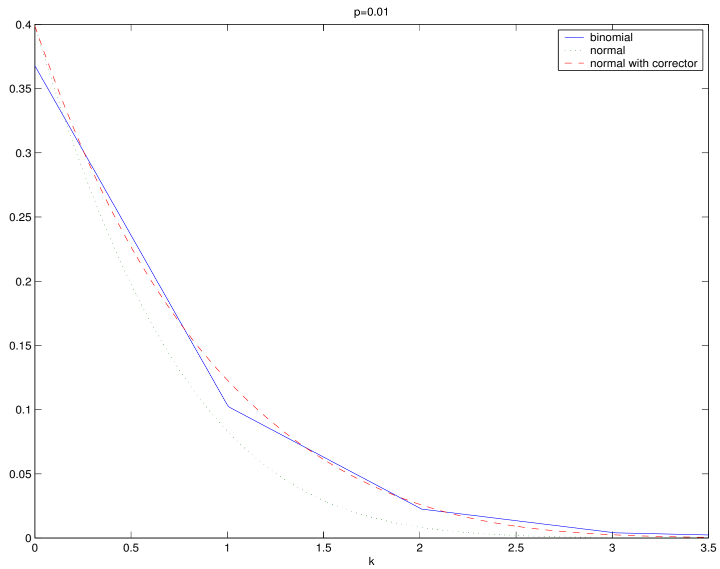
   (b) its normal approximation $\Phi_1((x-k)^+) = \mathbb{E}[(Z-k)^+]$ where Z is the standard normal random variable. Or $\phi_1(k) - k(1-\Phi_1(k))$ explicitly;

   (c) the corrected normal approximation $\Phi_1((x-k)^+) + C_{(x-k)^+}$.

   The expectation is presented as a function of the strike k. We fix the parameter $n = 100$ and we compare different values of p. In Figure 3.1, $p = 0.01$ and in Figure 3.2, $p = 0.1$. We remark that the binomial curve is piecewise linear because of the discretization. The length between two discretization points is γ, where $\gamma \approx 1$ in the first graph and $\gamma = \frac{1}{3}$ in the second graph. When the value of p is larger, the normal approximation becomes more robust, which corresponds to the common rule that when $np > 10$, we can apply the normal approximation to the binomial law.

   Both graphs show that our correction is effective. In Figure 3.1, it is obvious that the corrected curve fits better the piecewise binomial curve than the normal approximation curve without correction. In Figure 3.2, the corrected curve and the binomial curve almost coincide each other. At last, we note that at the point $k = 0$, the two approximative curves meet since there is no correction.

Figure 3.1: Homogenous call: $n = 100$, $p = 0.01$. Here $np = 1$, the correction is significant.



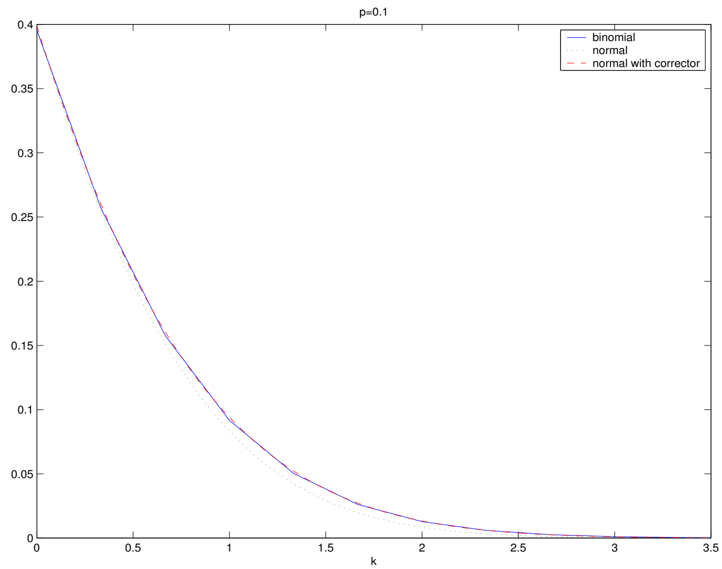
2. Call function: the exogenous case, (Figure 3.3 and 3.4). We repeat the above test for the exogenous case where X_i are independent but non-identically distributed. The only difference lies in the calculation of $\mathbb{E}[X_I^*]$. We simulate X_i as follows. Let Y_i be standard $0-1$ Bernoulli random variables of parameter p_i and let $X_i = \frac{Y_i - p_i}{\sqrt{\sum_{l=1}^n p_l(1-p_l)}}$. Then $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}[X_i] = \frac{p_i(1-p_i)}{\sum_{l=1}^n p_l(1-p_l)}$ and thus the sum variable $W = \sum_{i=1}^n X_i$ is of expectation zero and of variance 1. We fix $n = 100$ as above.

   For the first graph, we take $p_i = 0.02 \times U_i$ where U_i are independent uniform random variables on $[0,1]$. So the mean value of p is equal to 0.01. For the second graph, we regroup the $n = 100$ random variables into 10 groups and let all random variables in one group take a same value of p_i. In addition, we take 10 equally spaced values of p_i from 0.055 to 0.145 so that their mean value equals $\bar{p} = 0.1$.

   Figure 3.3 and Figure 3.4 resembles respectively Figure 3.1 and Figure 3.2 in the first test. We observe that the correction is notably related to the mean value of the probability p.

3. Call function: the asymptotic behavior, (Figure 3.5, 3.6 and 3.7). In this test, we fix the strike value k and we are interested in the asymptotic property concerning the parameter n. We take i.i.d. random variables X_i and let $\sigma_W = \mathrm{Var}[W] = 1$.

Figure 3.2: Homogenous call: $n = 100$ and $p = 0.10$. The binomial curve is rather near the normal curve, however, with our correction, the two curves coincide.

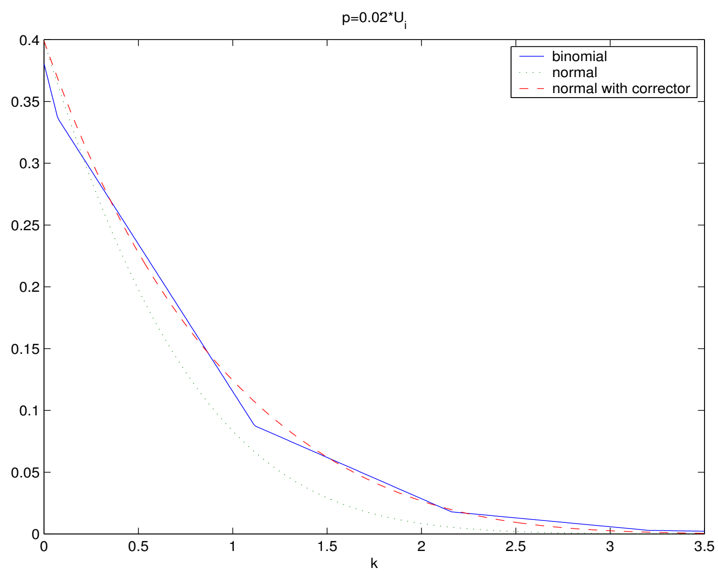


The strike k is set to be 1. The normal approximation in this case is a constant. Figure 3.5 shows the symmetric case where $p = \frac{1}{2}$. As been pointed out in [25], we observe oscillations in the binomial curve as a function of the size n of the portfolio. (In [25], n is the number of the time steps.) When n tends to infinity, the binomial curve converges to its normal approximation (the Black-Scholes price in [25]) which is presented as the horizontal line. The correction in this case is zero, so the corrected curve coincides with the normal curve without correction. Figure 3.6 and 3.7 show the asymmetric case where $p = 0.01$ in 3.6 and $p = 0.1$ in 3.7. In both graphs, we still observe oscillations in the binomial curves. When the value of p is smaller, the oscillation is slower. In the asymmetric case, there is a gap between the binomial and the normal curves with the same number of n, which means that the convergence speed is much slower. However, the binomial curves oscillate around the corrected curves in the two graphs, which means that our corrector is very efficient in the asymmetric case. Note that the corrected curve is situated in the upper part of the binomial curve, which almost serves an upper envelop of the binomial approximation.

3.4.3.2 The indicator function

The indicator function $h(x) = \mathbb{1}_{\{x>k\}}$ is less regular than the call function. Its derivative is a Dirac measure in the distribution sense. We can calculate the corrector C_h

Figure 3.3: Exogenous call: n=100, $p_i = 0.02 \times U_i$, so $\bar{p} = 0.01$.



using Corollary 3.4.10 and get

$$\begin{aligned} C_h &= \frac{\mathbb{E}(X_1^*)}{3}\left[\frac{1}{\sigma^2}\Phi_\sigma(x^2h'(x)) - \Phi_\sigma(h')\right] \\ &= \frac{(1-2p)}{6\sqrt{np(1-p)}}\left[\frac{1}{\sigma^2}k^2\phi_\sigma(k) - \phi_\sigma(k)\right]. \end{aligned}$$

However, we have no estimation for the corrected approximation. The zero correction points are $k = \sigma$ or $k = -\sigma$. Lacking theoretical result, we shall present numerically the correction effect.

Figure 3.8 and Figure 3.9 show the tests. In each graph, the reported quantity is the probability $\mathbb{P}(W > k)$ and its normal approximation with and without the correction term. The probability is presented as a function of the size number n. So we see the asymptotic behavior and we compare different values of p and k in the four graphs. In Figure 3.8, $p = 0.01$ and in Figure 3.9, $p = 0.1$. The strike values are $k = 0$ and $k = 3$. When $k = 3$, the two graphs show that our correction is effective, while when $k = 0$, it is hard to tell whether the correction improves the approximation quality because of the discretization effet.

It is well-known that the convergence speed of the indicator function is of order $O(\frac{1}{\sqrt{n}})$ by the Berry-Esseen inequality. However, with these numerical result, we have naturally the intuition that after our correction, there exist some non-uniform estimations of the convergence speed according to the values of k. We can not yet explain

Figure 3.4: Exogenous call: $n = 100$ and there are 10 random variables which take the same value of p_i. We take 10 value of p_i from 0.055 to 0.145 equally dispersed, so that $\bar{p} = 0.1$.

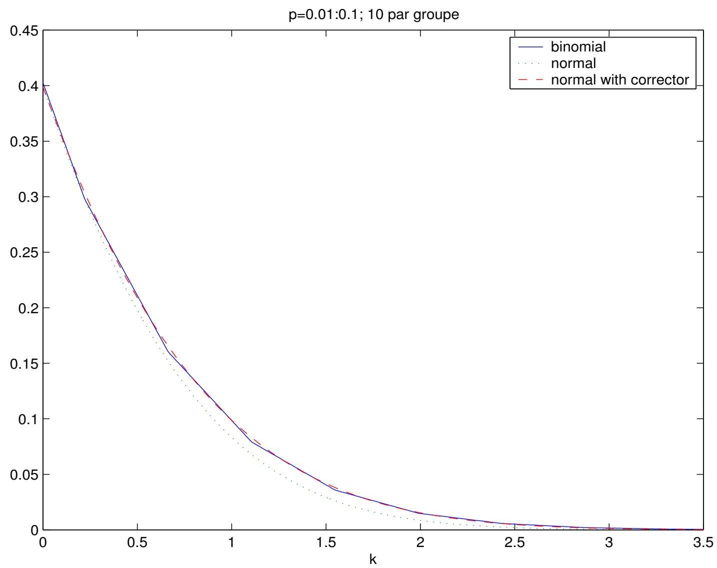


Figure 3.5: Asymptotic call: $k = 1$ and $p = 0.5$.

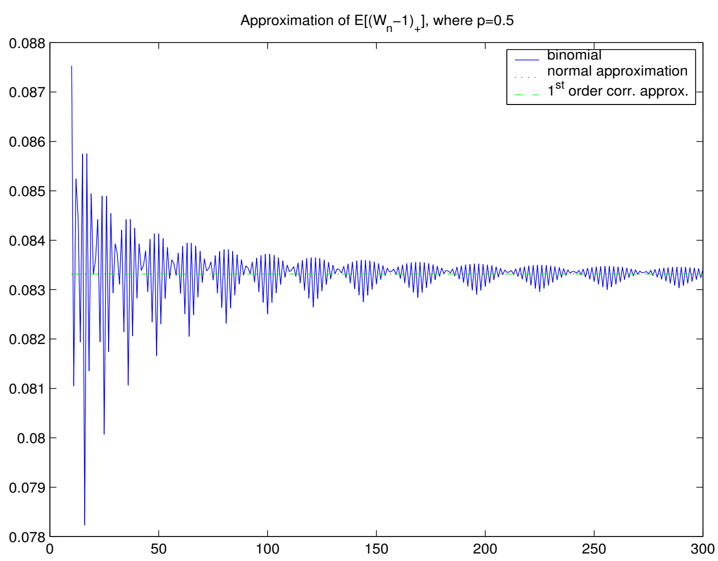


Figure 3.6: Asymptotic call: $k = 1$ and $p = 0.01$.

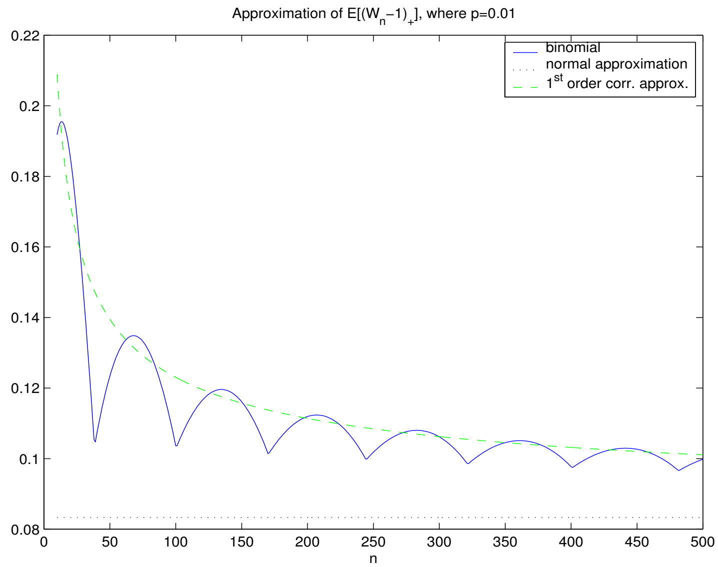
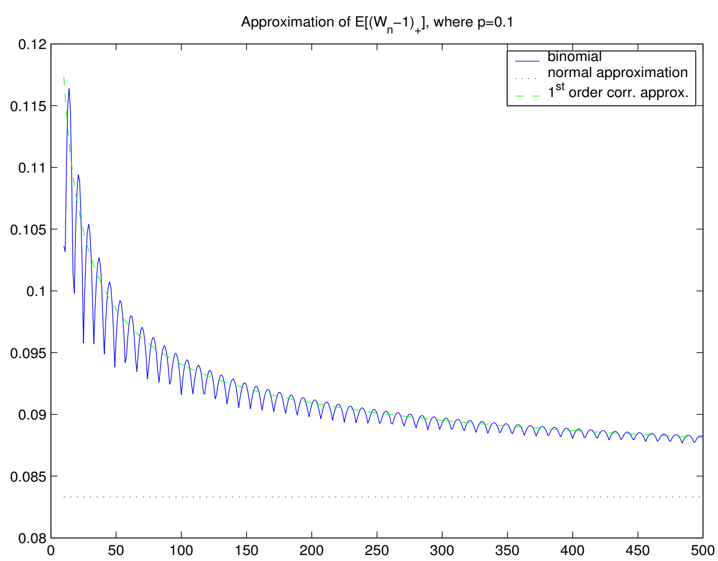
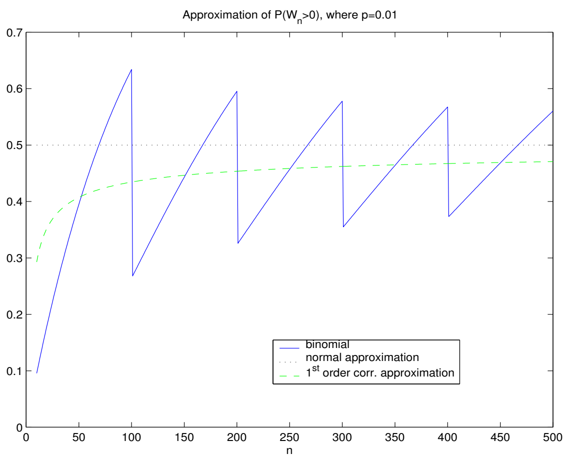


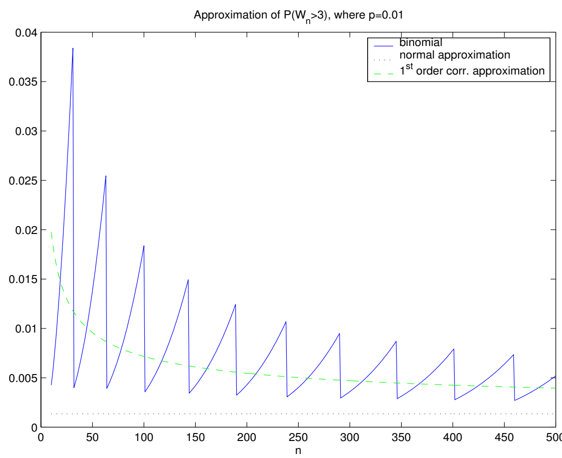
Figure 3.7: Asymptotic call: $k = 1$ and $p = 0.1$.



this phenomenon with the previous discussion. However, we think it's worth further study to well understand this problem.

Figure 3.8: Indicator function or the probability function: $p = 0.01$. The two graphs are for $k = 0$ and $k = 3$.





3.4.4 Saddle-point method

In Antonov, Mechkov and Misirpashaev [2], the authors propose efficient correction terms to calculate the conditional expectation $\mathbb{E}[(W-k)^+]$ using the saddle point method. This subsection concentrates on this method. We first introduce their results and we then interpret the main idea from a more probabilistic point of view. We show that choosing a saddle point can be viewed as a change of probability and we apply our correction of Theorem 3.4.8 under the new probability measure. We shall compare the results obtained by our method and by the method in their paper through some numerical tests.

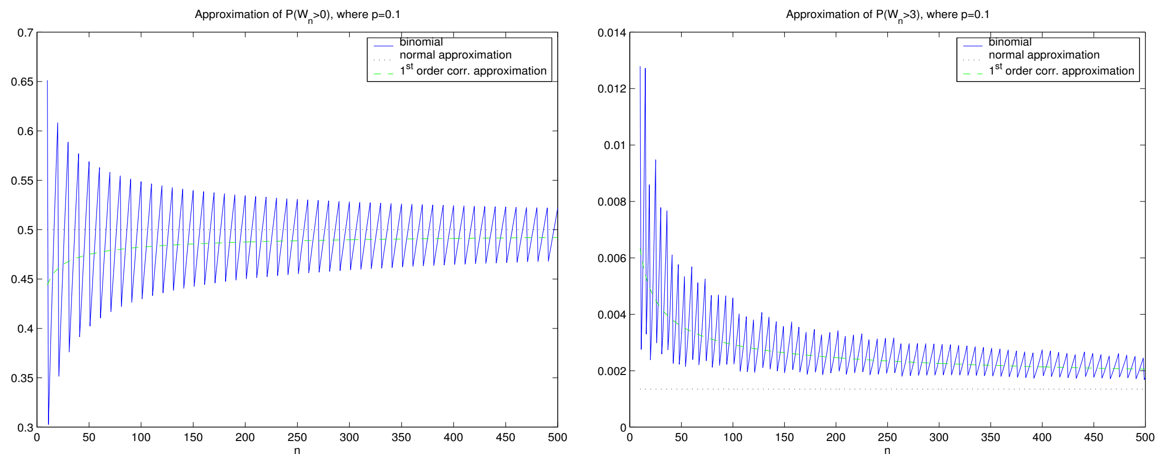
Our objective is to calculate some expectation functions under a given probability $\mathbb{P}$. The saddle-point method consists of writing the concerned functions as an integral of some function of the cumulant generating function $\mathcal{K}(\xi) = \ln \mathbb{E}_{\mathbb{P}}[\exp(\xi W)]$. For example,

$$\mathbb{E}_{\mathbb{P}}[(W-k)^+] = \lim_{A\to+\infty} \frac{1}{2\pi i} \int_{c-iA}^{c+iA} \frac{\exp(\mathcal{K}(\xi)-\xi k)}{\xi^2} d\xi \tag{3.63}$$

and

$$\mathbb{P}(W \geq k) = \lim_{A\to+\infty} \frac{1}{2\pi i} \int_{c-iA}^{c+iA} \frac{\exp(\mathcal{K}(\xi)-\xi k)}{\xi} d\xi.$$

Figure 3.9: Indicator function: $p = 0.1$, $k = 0$ and $k = 3$.



In general, the expansion of the integrand function is made around its critical point, the saddle point, where the integrand function decreases rapidly and hence is most dense. In [2], the saddle point is chosen such that

$$\mathcal{K}'(\xi_0) = k.$$

Making expansion of $\exp(\mathcal{K}(\xi) - \xi k)$ around the saddle point ξ_0 in (3.63), the authors propose to approximate $\mathbb{E}_{\mathbb{P}}[(W-k)^+]$ by the following terms which are of increasing precision orders:

1. $C_0^s = (\mathbb{E}[W] - k)^+$,
2. $C_1^s = C_0^s + e^{\mathcal{K}(\xi_0) - \xi_0 k} J_2\big(\mathcal{K}''(\xi_0), \xi_0\big)$,
3. $C_2^s = C_1^s + \frac{1}{6}\xi_0 \mathcal{K}^{(3)}(\xi_0) e^{\mathcal{K}(\xi_0) - \xi_0 k}$
   $\times \Big(-2J_0\big(\mathcal{K}''(\xi_0), \xi_0\big) + 3\xi_0 J_1\big(\mathcal{K}''(\xi_0), \xi_0\big) - \xi_0^2 J_2\big(\mathcal{K}''(\xi_0), \xi_0\big)\Big)$.

where

$$\begin{cases} J_0(m,\xi) = \frac{1}{\sqrt{2\pi m}}, \\ J_1(m,\xi) = \mathrm{sign}(\xi_0) e^{\frac{1}{2} m \xi_0^2} \mathcal{N}_1(-m|\xi_0|), \\ J_2(m,\xi) = \sqrt{\frac{m}{2\pi}} - m|\xi_0| e^{\frac{1}{2} m \xi_0^2} \mathcal{N}_1(-m|\xi_0|). \end{cases}$$

Although numerical results show that these approximations are efficient compared to the standard normal approximation, no theoretical discussion concerning the estimation errors is mentioned in their paper.

In the saddle point method, the first step is to find the value of the saddle point ξ_0. This is equivalent to solve the equation $\mathcal{K}'(\xi) = k$, which can be rather tedious sometimes. In the homogeneous case where X_i follow i.i.d. asymmetric Bernoulli distributions, we have the explicit solution. Since X_i are identically distributed, then $\gamma_1 = \cdots = \gamma_n = \frac{\sigma_W}{\sqrt{npq}}$. Hence

$$\mathcal{K}'(\xi) = n\frac{\mathbb{E}_{\mathbb{P}}[X_1 e^{\xi X_1}]}{\mathbb{E}_{\mathbb{P}}[e^{\xi X_1}]} = \frac{\sigma_W\sqrt{npq}(e^{\xi_0\gamma} - 1)}{pe^{\xi_0\gamma} + q} = k.$$

This equation has a unique solution

$$\xi_0 = \frac{\sqrt{npq}}{\sigma_W}\ln\Big(\frac{\sigma_W\sqrt{npq} + kq}{\sigma_W\sqrt{npq} - kp}\Big).$$

In the exogenous case, the cumulant generating function $\mathcal{K}(\xi)$ is calculated by

$$\mathcal{K}(\xi) = \ln\mathbb{E}_{\mathbb{P}}[e^{\xi W}] = \sum_{i=1}^{n}\ln\mathbb{E}_{\mathbb{P}}[e^{\xi X_i}] = \sum_{i=1}^{n}\ln(pe^{\xi\gamma_i q} + qe^{-\xi\gamma_i p}).$$

As a consequence,

$$\mathcal{K}'(\xi) = \sum_{i=1}^{n}\frac{pq\gamma_i(e^{\xi\gamma_i} - 1)}{pe^{\xi\gamma_i} + q}, \quad \mathcal{K}''(\xi) = \sum_{i=1}^{n}\frac{pq\gamma_i^2 e^{\xi\gamma_i}}{(pe^{\xi\gamma_i} + q)^2}.$$

Hence, we can obtain ξ_0 by solving numerically $\sum_{i=1}^{n}\frac{pq\gamma_i(e^{\xi\gamma_i}-1)}{pe^{\xi\gamma_i}+q} = k$. And then the approximations C_1^s and C_2^s can be obtained. Compared to our method, the saddle point method demands more calculation to get the correction terms, in the homogenous case and especially in the exogenous case.

Figure 3.10 and 3.11 compare the correction results in the homogenous case by different methods:

1) the normal approximation,

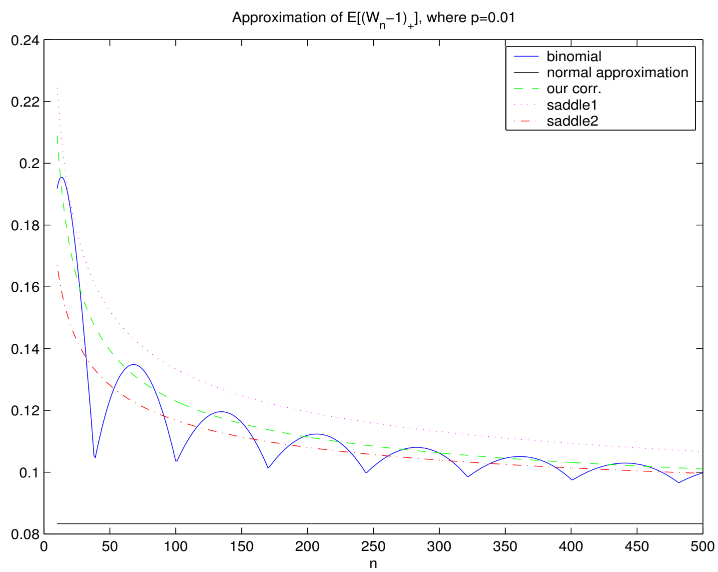
2) the normal approximation with our correction,

3) the saddle point method with the first correction C_1^s of [2],

4) the saddle point method with the second correction C_2^s of [2].

We repeat the test which produce the Figure 3.6 and 3.7. The strike is fixed to be $k = 1$ and the asymptotic behavior of the correction is showed for $p = 0.01$ and $p = 0.1$. We observe that our correction is better than the first order correction but is less effective than the second order correction of the saddle point method.

Figure 3.10: Saddle point comparison, asymptotic call: $p = 0.01$.



3.4.4.1 A probabilistic point of view

We now interpret the saddle point method from a more probabilistic point of view. First, let us choose the saddle point ξ_0 such the $\mathcal{K}'(\xi_0) = k$. Second, if we let $\mathbb{Q}$ to be an equivalent probability measure such that

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{\exp(\xi_0 W)}{\mathbb{E}[\exp(\xi_0 W)]},$$

then

$$\mathbb{E}_{\mathbb{Q}}[W] = \frac{\mathbb{E}_{\mathbb{P}}[We^{\xi_0 W}]}{\mathbb{E}_{\mathbb{P}}[e^{\xi_0 W}]} = \mathcal{K}'(\xi_0) = k. \tag{3.64}$$

That is to say, under the probability measure $\mathbb{Q}$, the expectation of the sum variable W equals the strike k. Now, consider the expectation of the call function, we have

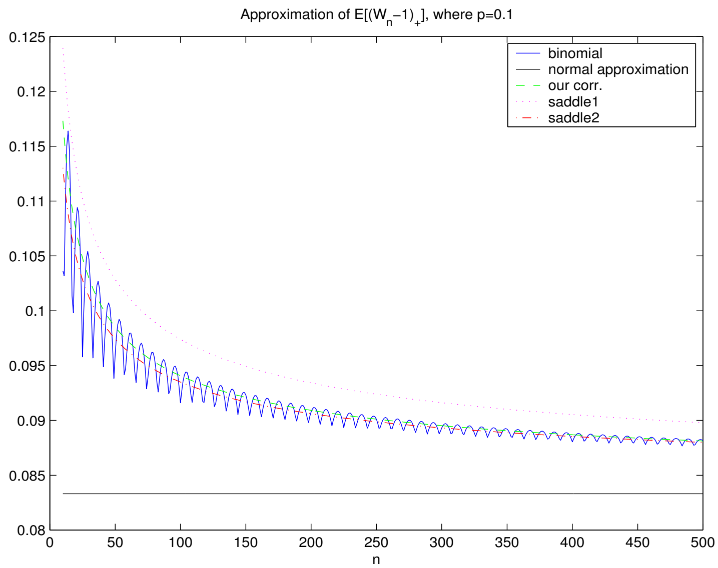
$$\mathbb{E}_{\mathbb{P}}[(W-k)^+] = \mathbb{E}_{\mathbb{Q}}[e^{-\xi_0 W + \mathcal{K}(\xi_0)}(W-k)^+].$$

Note that W is no longer of expectation zero under the probability $\mathbb{Q}$. To apply our previous result, we need to centralize the random variable. Let $\widehat{W} = W - k$, then the above equality can be written as

$$\mathbb{E}_{\mathbb{P}}[(W-k)^+] = e^{-\xi_0 k + \mathcal{K}(\xi_0)}\mathbb{E}_{\mathbb{Q}}[e^{-\xi_0 \widehat{W}}\widehat{W}^+]. \tag{3.65}$$

Since $\mathbb{E}_{\mathbb{Q}}[\widehat{W}] = 0$, now the problem is to approximate the function $e^{-\xi_0 \widehat{W}}\widehat{W}^+$ under the probability $\mathbb{Q}$, where $\widehat{W}$ is a zero-mean random variable which can be written as the

Figure 3.11: Saddle point comparison, asymptotic call: $p = 0.1$.



sum of independent asymmetric Bernoulli random variables, i.e. $\widehat{W} = \widehat{X}_1 + \cdots + \widehat{X}_n$ where $\widehat{X}_i = X_i - \mathbb{E}_{\mathbb{Q}}[X_i]$.

Example 3.4.17 We revisit the example of the asymmetric Bernoulli random variables. Under the original probability $\mathbb{P}$, $X_i \sim B_{\gamma_i}(q, -p)$, i.e. $\mathbb{P}(X_i = \gamma_i q) = p$ and $\mathbb{P}(X_i = -\gamma_i p) = q$ where $\gamma_i = \frac{\sigma_i}{\sqrt{p(1-p)}}$. By the change of probability,

$$\begin{aligned}\mathbb{Q}(X_i = \gamma_i q) &= \mathbb{E}_{\mathbb{P}}\Big[1\!\!1_{\{X_i=\gamma_i q\}} \frac{\exp(\xi_0 W)}{\mathbb{E}[\exp(\xi_0 W)]}\Big] \\ &= \frac{\mathbb{E}_{\mathbb{P}}[1\!\!1_{\{X_i=\gamma_i q\}} \exp(\xi_0 X_i)]}{\mathbb{E}_{\mathbb{P}}[\exp(\xi_0 X_i)]} = \frac{p}{p + e^{-\xi_0\gamma_i} q},\end{aligned}$$

Similarly, $Q(X_i = -\gamma_i p) = \frac{q}{pe^{\xi_0\gamma_i} + q}$ and then

$$\mathbb{E}_{\mathbb{Q}}[X_i] = \frac{\gamma_i pq(e^{\xi_0\gamma_i} - 1)}{pe^{\xi_0\gamma_i} + q}.$$

The centralized random variables $\widehat{X}_i$ are zero-mean asymmetric Bernoulli random variables which satisfy

$$\mathbb{Q}\Big(\widehat{X}_i = \frac{\gamma_i q}{pe^{\xi_0\gamma_i} + q}\Big) = \frac{p}{p + qe^{-\xi_0\gamma_i}}, \quad \mathbb{Q}\Big(\widehat{X}_i = -\frac{\gamma_i pe^{\xi_0\gamma_i}}{pe^{\xi_0\gamma_i} + q}\Big) = \frac{q}{pe^{\xi_0\gamma_i} + q}.$$

We now apply Theorem 3.4.8 to calculate $\mathbb{E}_{\mathbb{Q}}[e^{-\xi_0 \widehat{W}}\widehat{W}^+]$. This is a function of similar property with the call function in the sense that its derivative contains an indicator function and its second order derivative exists in the distribution sense. We now give the approximation formula and the correction term.

Proposition 3.4.18 *Let $h_{\xi_0} = e^{-\xi_0 x}x^+$. Then the normal approximation of the expectation $\mathbb{E}_{\mathbb{Q}}[h_{\xi_0}(\widehat{W})]$ is given by*

$$\mathbb{E}_{\mathbb{Q}}[e^{-\xi_0 Z}Z^+] = e^{\frac{1}{2}\xi_0^2\widehat{\sigma}^2}\Big(\widehat{\mu}\big(1-\mathcal{N}_1(-\frac{\widehat{\mu}}{\widehat{\sigma}})\big)+\widehat{\sigma}\phi_1\big(-\frac{\widehat{\mu}}{\widehat{\sigma}}\big)\Big) \tag{3.66}$$

where $Z \sim N(0,\widehat{\sigma}^2)$, $\widehat{\mu} = -\xi_0\mathcal{K}''(\xi_0)$ and $\widehat{\sigma}^2 = \mathcal{K}''(\xi_0)$. Moreover, the correction term is given by

$$C_{h_{\xi_0}} = \frac{1}{\widehat{\sigma}^2}\mathbb{E}_{\mathbb{Q}}[\widehat{X}_I^*]\mathbb{E}_{\mathbb{Q}}\Big[e^{-\xi_0 Z}\Big(\frac{(Z^+)^4}{3\widehat{\sigma}^2}-(Z^+)^2\Big)\Big] \tag{3.67}$$

$$= e^{\frac{1}{2}\xi_0^2\widehat{\sigma}^2}\frac{\mathbb{E}_{\mathbb{Q}}[\widehat{X}_I^*]}{\widehat{\sigma}^2}\bigg(\Big(\frac{\widehat{\mu}^4}{3\widehat{\sigma}^2}+\widehat{\mu}^2\Big)\Big(1-\mathcal{N}_1\big(-\frac{\widehat{\mu}}{\widehat{\sigma}}\big)\Big)+\Big(\frac{\widehat{\mu}^3}{3\widehat{\sigma}}+\frac{2\widehat{\mu}\widehat{\sigma}}{3}\Big)\phi_1\big(-\frac{\widehat{\mu}}{\widehat{\sigma}}\big)\bigg) \tag{3.68}$$

Proof. First, we verify that $\mathrm{Var}_{\mathbb{Q}}[W] = \mathcal{K}''(\xi_0)$. In fact,

$$\mathrm{Var}_{\mathbb{Q}}[W] = \mathbb{E}_{\mathbb{Q}}[W^2]-\mathbb{E}_{\mathbb{Q}}[W]^2 = \frac{\mathbb{E}_{\mathbb{P}}[W^2e^{\xi_0 W}]}{\mathbb{E}_{\mathbb{P}}[e^{\xi_0 W}]}-\bigg(\frac{\mathbb{E}_{\mathbb{P}}[We^{\xi_0 W}]}{\mathbb{E}_{\mathbb{P}}[e^{\xi_0 W}]}\bigg)^2$$
$$= \bigg(\frac{\mathbb{E}_{\mathbb{P}}[We^{\xi_0 W}]}{\mathbb{E}_{\mathbb{P}}[e^{\xi_0 W}]}\bigg)' = \mathcal{K}''(\xi_0).$$

Then the normal approximation is $\Phi_{\widehat{\sigma}}(h_{\xi_0}) = \mathbb{E}_{\mathbb{Q}}[e^{-\xi_0 Z}Z^+]$ and the corrector (3.67) is obtained by Theorem 3.4.8. The last step consists of calculating explicitly the approximation and the correction terms. To this end, we introduce another change of probability to simplify the computation. Let

$$\frac{d\mathbb{P}_0}{d\mathbb{Q}} = e^{-\xi_0 Z-\frac{1}{2}\xi_0^2\mathcal{K}''(\xi_0)}.$$

Then

$$\mathbb{E}_{\mathbb{Q}}[e^{-\xi_0 Z}f(Z)] = \mathbb{E}_{\mathbb{P}_0}\big[e^{-\xi_0 Z}f(Z)\frac{d\mathbb{Q}}{d\mathbb{P}_0}\big] = e^{\frac{1}{2}\xi_0^2\mathcal{K}''(\xi_0)}\mathbb{E}_{\mathbb{P}_0}[f(Z)]$$

for any function f such that the quantities of the two sides are well defined. To obtain (3.66), let $f(x) = x^+$. Notice that under the probability $\mathbb{P}_0$, Z is still a normal random variable with $\widehat{\mu} = \mathbb{E}_{\mathbb{P}_0}[Z] = -\xi_0\mathcal{K}''(\xi_0)$ and $\mathrm{Var}_{\mathbb{P}_0}[Z] = \mathcal{K}''(\xi_0) = \widehat{\sigma}^2$. Then

$$\mathbb{E}_{\mathbb{P}_0}[Z^+] = \mathbb{E}_{\mathbb{P}_0}[\mathbb{1}_{\{\widehat{\mu}+\widehat{\sigma}Z_0\geq 0\}}(\widehat{\mu}+\widehat{\sigma}Z_0)]$$
$$= \widehat{\mu}\Big(1-\mathcal{N}_1\big(-\frac{\widehat{\mu}}{\widehat{\sigma}}\big)\Big)+\widehat{\sigma}\phi_1\big(-\frac{\widehat{\mu}}{\widehat{\sigma}}\big)$$

where $Z_0 \sim N(0,1)$ and (3.66) follows immediately. To obtain (3.68), consider respectively $\mathbb{E}_{\mathbb{P}_0}[(Z^+)^2]$ and $\mathbb{E}_{\mathbb{P}_0}[(Z^+)^4]$. Combining the invariant property (3.7), we get similarly as above

$$\mathbb{E}_{\mathbb{P}_0}[(Z^+)^2] = (\widehat{\mu}^2 + \widehat{\sigma}^2)\Big(1 - \mathcal{N}_1\big(-\frac{\widehat{\mu}}{\widehat{\sigma}}\big)\Big) + \widehat{\mu}\widehat{\sigma}\phi_1\big(-\frac{\widehat{\mu}}{\widehat{\sigma}}\big)$$

and

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_0}[(Z^+)^4] &= \widehat{\mu}\mathbb{E}_{\mathbb{P}_0}[\mathbb{1}_{\{\widehat{\mu}+\widehat{\sigma}Z_0\geq 0\}}(\widehat{\mu}+\widehat{\sigma}Z_0)^3] + \widehat{\sigma}\mathbb{E}_{\mathbb{P}_0}[Z_0\mathbb{1}_{\{\widehat{\mu}+\widehat{\sigma}Z_0\geq 0\}}(\widehat{\mu}+\widehat{\sigma}Z_0)^3] \\ &= \widehat{\mu}^2\mathbb{E}_{\mathbb{P}_0}[\mathbb{1}_{\{\widehat{\mu}+\widehat{\sigma}Z_0\geq 0\}}(\widehat{\mu}+\widehat{\sigma}Z_0)^2] + \widehat{\mu}\widehat{\sigma}\mathbb{E}_{\mathbb{P}_0}[Z_0\mathbb{1}_{\{\widehat{\mu}+\widehat{\sigma}Z_0\geq 0\}}(\widehat{\mu}+\widehat{\sigma}Z_0)^2] \\ &+ \widehat{\sigma}\mathbb{E}_{\mathbb{P}_0}[(\mathbb{1}_{\{\widehat{\mu}+\widehat{\sigma}Z_0\geq 0\}}(\widehat{\mu}+\widehat{\sigma}Z_0)^3)'] \\ &= \widehat{\mu}^2\mathbb{E}_{\mathbb{P}_0}\mathbb{E}[(Z^+)^2] + \widehat{\mu}\widehat{\sigma}\mathbb{E}[(\mathbb{1}_{\{\widehat{\mu}+\widehat{\sigma}Z_0\geq 0\}}(\widehat{\mu}+\widehat{\sigma}Z_0)^2)'] + 3\widehat{\sigma}^2\mathbb{E}[(Z^+)^2] \\ &= (\widehat{\mu}^2 + 3\widehat{\sigma}^2)\mathbb{E}_{\mathbb{P}_0}[(Z^+)^2] + 2\widehat{\mu}\widehat{\sigma}^2\mathbb{E}_{\mathbb{P}_0}[Z^+],\end{aligned}$$

which deduces (3.68). □

3.5 Application to CDOs portfolios

In this section, we apply the results of the previous section to the evaluation of a CDO tranche. As mentioned before, we proceed in two steps. The approximation correction is used to calculate the expectation of the conditional cumulative losses. In the second step, we integrate the conditional losses function with respect to the density function of Y and we study the correlation parameter ρ. Two points to be noted are

1) the normalization of the standard $0-1$ Bernoulli random variables to the zero-mean asymmetric Bernoulli random variables.

2) the probability p is now a function of the common factor Y. We shall see that the form of $p(Y)$ has an impact on the correction.

In the following of this section, we are under the factor model framework as introduced in Subsection 3.1.1.

3.5.1 Conditional loss approximation

Consider a CDO portfolio of n credits where the weight of each firm is denoted by $w_i = \frac{N_i}{N}$ and $\sum_{i=1}^n w_i = 1$. The percentage loss is given by

$$l_T = \sum_{i=1}^n w_i(1-R_i)\mathbb{1}_{\{\tau_i \leq T\}}. \tag{3.69}$$

We are interested in the conditional expectation $\mathbb{E}[(l_T - K)^+|Y]$ where K is the tranche threshold and Y is the common factor. For synthetic CDOs, the threshold values are fixed to be $K = 3\%, 6\%, 9\%, 12\%$.

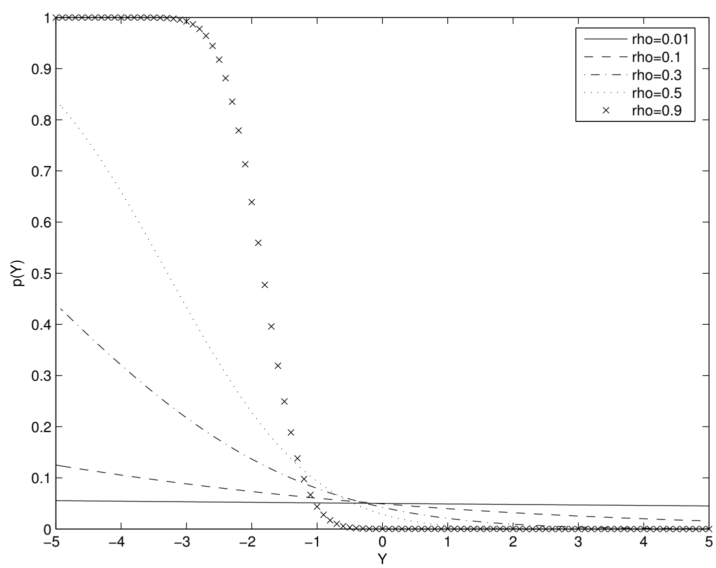
Denote by $\xi_i = 1\!\!1_{\{\tau_i \leq T\}}$ to be the indicator function of default for each credit. They are standard $0-1$ Bernoulli random variables. Moreover, conditional on the factor Y, they are independent and of probability parameters $p_i(T|Y)$. In the normal factor case, $p_i(T|Y)$ is given by (3.2), i.e. $p_i(T|Y) = \mathcal{N}\big(\frac{\mathcal{N}^{-1}(\alpha_i) - \sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\big)$ where $\alpha_i = 1 - q_i(T)$ is the expected default probability of the credit i before the maturity T. In fact, it's easy to verify that

$$\mathbb{E}[p_i(T|Y)] = \mathbb{P}(\tau_i \leq T) = \alpha_i.$$

To simplify the notation, in this subsection, we write $p_i(Y)$ instead of $p_i(T|Y)$. In addition, it is often supposed that ρ_i are identical, in this case, $p_i(Y) = p(Y) = \mathcal{N}\big(\frac{\mathcal{N}^{-1}(\alpha) - \sqrt{\rho}Y}{\sqrt{1-\rho}}\big)$.

Figure 3.12 shows $p(Y)$ as a function of the factor Y which is decreasing. We compare different values of ρ and we notice that this correlation parameter plays a significant role. When ρ approaches zero, $p(Y)$ converges to a constant which equals α. When ρ increases, the values of $p(Y)$ disperse in $[0, 1]$ and when ρ tends to 1, we observe a concentration of $p(Y)$ at two values 0 and 1. For simplicity, in the following of this subsection, we write p instead of $p(Y)$ since we only consider the conditional case.

Figure 3.12: The function $p(Y)$: $\alpha = 0.05$. Correlation parameters are $\rho = 0.01$, 0.1, 0.3, 0.5 and 0.9.



To apply the approximation correction, we first normalize the Bernoulli random variables to be of expectation zero. In addition, the sum variable should be of finite variance. We denote the variance of $\omega_i(1-R_i)\mathbb{1}_{\{\tau_i\le T\}}$ conditioned on Y by $\sigma_i^2 = \omega_i^2(1-R_i)^2 p_i(1-p_i)$ and let $\Sigma = \mathrm{Var}[l_T|Y] = \sqrt{\sum_{i=1}^n \sigma_i^2}$. Let

$$X_i = \frac{\sigma_i}{\Sigma}\frac{\xi_i - p_i}{\sqrt{p_i(1-p_i)}},$$

then $\mathbb{E}[X_i]=0$ and $\mathrm{Var}[X_i]=\sigma_i^2/\Sigma^2$. Conditioned on Y, X_i are zero-mean asymmetric Bernoulli random variables such that $X_i \sim \mathcal{B}_{\gamma_i}(q_i,p_i)$ with $\gamma_i = \frac{\sigma_i}{\Sigma}\frac{1}{\sqrt{p_i(1-p_i)}}$. Denote their sum by $W=\sum_{i=1}^n X_i$, then $\mathbb{E}[W]=0$ and $\mathrm{Var}[W]=1$. The percentage loss can thus be written as $l_T = \Sigma W + \sum_{i=1}^n \omega_i(1-R_i)p_i$ and

$$(l_T-K)^+ = \Sigma\left(W - \frac{K-\sum_{i=1}^n \omega_i(1-R_i)p_i}{\Sigma}\right)^+. \tag{3.70}$$

Note that the strike $K_Y = \big(K-\sum_{i=1}^n \omega_i(1-R_i)p_i\big)/\Sigma$ is a function of p_i and eventually of Y. We can now apply the correction (3.57) to the expectation of (3.70). So the correction to the normal approximation of $\mathbb{E}\big[(l_T-K)^+|Y\big]$ is given by

$$\frac{1}{3}\Sigma\mathbb{E}[X_I^*|Y]K_Y\phi_1(K_Y) = \frac{1}{6}\sum_{i=1}^n \frac{\sigma_i^3}{\Sigma^2}\frac{1-2p_i}{\sqrt{p_i(1-p_i)}}K_Y\phi_1(K_Y) \tag{3.71}$$

where p_i, σ_i, Σ and K_Y are all functions of Y.

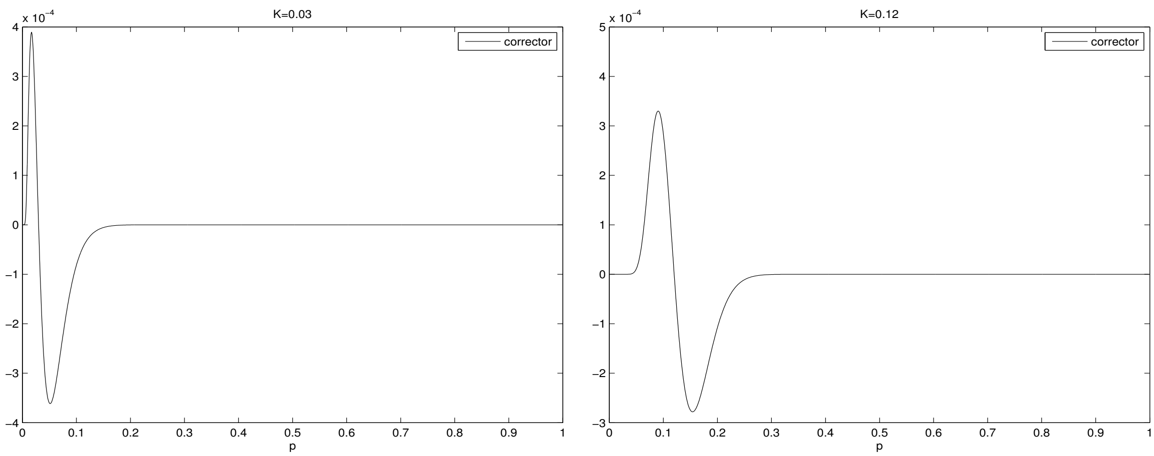
In the homogeneous portfolio case where all parameters of each credit are supposed to be equal, we have $\omega_i=\frac{1}{n}$, $R_i=R$ and $p_i=p$. Then $K_Y = \frac{(K-(1-R)p)\sqrt{n}}{(1-R)\sqrt{p(1-p)}}$. The corrector of $\mathbb{E}[(l_T-K)^+|Y]$ is given by

$$\frac{1-R}{6n}(1-2p)K_Y\phi_1(K_Y). \tag{3.72}$$

Figure 3.13 show the corrector (3.72) as a function of p when the strike value K is fixed. The corrector vanishes when $K_Y=0$, which means that the strike K equals the expected loss, i.e. $K=\mathbb{E}[l_T|Y]=(1-R)p$. Moreover, the function $x\phi_1(x)$ attains its maximum and minimum values at $x=1$ and $x=-1$ respectively, that is $K=(1-R)\big(p\pm\sqrt{\frac{p(1-p)}{n}}\big)$, and converges rapidly to zero when the absolute value of x increases. This is shown by the two graphes where we suppose $R=0$. The corrector equals zero when $p=K$ and attains the extreme values when p is rather near K and then vanishes rapidly when p moves away from K.

We then compare the approximation results of the conditional expectation of the call function $\mathbb{E}[(l_T-K)^+|Y]$ by the normal approximation, saddle point method and our method. The test is for the homogeneous case where $w_i=\frac{1}{n}$, $R_i=R=0$ and

Figure 3.13: Conditional Call corrector: $\alpha = 0.05$, $R = 0$, $n = 100$. $K = 0.03$ and $K = 0.12$ respectively.



$p_i = p$. In the conditional case, the probability p is viewed as a parameter, but not a function. The comparison baseline is calculated by the direct binomial simulation.

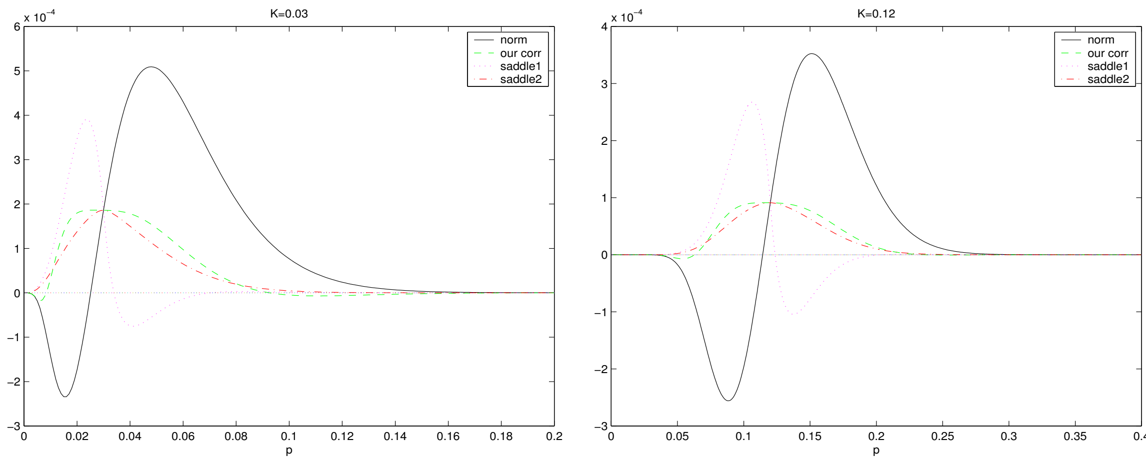
Figure 3.14 shows the approximation errors of $\mathbb{E}[(l_T - K)^+|Y]$ by different methods. The reported quantity is the difference between the values obtained by the approximation methods and the direct binomial simulation. For example, the curve denoted by normal approximation represents $\Phi((x-K)^+|Y) - \mathbb{E}[(l_T - K)^+|Y]$, so do the other curves. We fix the strike K and the horizontal axis is the probability p. We observe that both methods are effective compared to the direct normal approximation. All corrections are concentrated in a neighbourhood interval around the strike point K and when $p = K$, the correctors equal zero by the three methods. Our correction obtains the same order of precision with the second order correction of the saddle point method. We repeat however that the calculations in our case are much easier.

3.5.2 The common factor impact

In this subsection, we calculate the call expectation by integrating $\mathbb{E}[(l_T - K)^+|Y]$ with respect to the density function of Y. The approximation of $\mathbb{E}[(l_T - K)^+|Y]$ having been discussed previously, the expectation function can be easily obtained by integrating different approximations of the conditional expectation. We shall compare these results by numerical tests. In the following, we denote the corrector (3.71) by $C_l(Y)$. Since the probability $p(Y)$ is a function of the correlation parameter ρ, we are interested in its impact on the approximation result.

We observe from Figure 3.12 and Figure 3.13 that when there is light correlation,

Figure 3.14: Approximation error of conditional Call function: $n = 100$, $\alpha = 0.05$. $K = 0.03$ and $K = 0.12$ respectively.



that is, when ρ is small, the values of $p(Y)$ concentrate around the mean value α and the correction should be significant in this case. On the contrary, when there is strong correlation with large values of ρ, $p(Y)$ disperse in $[0, 1]$ and there should be little correction.

In the normal factor model, Y is a standard normal random variable. We can calculate the expectation on Y by changing the variable. For example, consider the corrector function, we have $\mathbb{E}[C_l(Y)] = \int_{-\infty}^{+\infty} C_l(y)d\mathcal{N}(y)$. A change of variable $z = p(y)$ yields

$$\mathbb{E}[C_l(Y)] = -\frac{1-R}{6n}\int_0^1 (1-2z)\frac{(K-z)\sqrt{n}}{\sqrt{z(1-z)}}\phi_1\Big(\frac{(K-z)\sqrt{n}}{\sqrt{z(1-z)}}\Big)\frac{\mathcal{N}'(p^{-1}(z))}{p'(p^{-1}(z))}dz.$$

Since $p'(y) = -\frac{\sqrt{\rho}}{\sqrt{1-\rho}}\mathcal{N}'\left(\mathcal{N}^{-1}(p(y))\right)$ and $p^{-1}(z) = \frac{\mathcal{N}^{-1}(\alpha)-\sqrt{1-\rho}\mathcal{N}^{-1}(z)}{\sqrt{\rho}}$, we have

$$\mathcal{N}'(p^{-1}(z)) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\mathcal{N}^{-1}(\alpha)-\sqrt{1-\rho}\mathcal{N}^{-1}(z))^2}{2\rho}\right),$$

$$p'(p^{-1}(z)) = -\frac{\sqrt{\rho}}{\sqrt{1-\rho}}\mathcal{N}'(\mathcal{N}^{-1}(z)) = -\frac{\sqrt{\rho}}{\sqrt{2\pi(1-\rho)}}\exp\left(-\frac{\mathcal{N}^{-1}(z)^2}{2}\right).$$

Then

$$\mathbb{E}[C_h(Y)] = \int_0^1 \frac{1}{6n}\sqrt{\frac{1-\rho}{\rho}}(1-2z)\frac{(K-z)\sqrt{n}}{\sqrt{z(1-z)}}\phi_1\Big(\frac{(K-z)\sqrt{n}}{\sqrt{z(1-z)}}\Big)\times$$
$$\exp\left(-\frac{1}{2\rho}\left(\mathcal{N}^{-1}(\alpha)^2+(1-2\rho)\mathcal{N}^{-1}(z)^2-2\sqrt{1-\rho}\mathcal{N}^{-1}(\alpha)\mathcal{N}^{-1}(z)\right)\right)dz. \tag{3.73}$$

Figure 3.15 and Figure 3.16 show the approximation error of $\mathbb{E}[(l_T - K)^+]$. For each curve, the reported quantity is the difference between values of the integrand function in (3.73) calculated by the direct binomial calculation and the approximation method. Therefore, the approximation error equals the integral of this difference on $[0, 1]$, which is represented by the area under the curve. So compared to Figure 3.14, we are not interested in the absolute value of the function at each point, but the whole area under each curve.

We compare different values of K and ρ in the following graphs. In Figure 3.15, $K = 0.03$, the two graphs are for $\rho = 0.2$ and $\rho = 0.8$. In Figure 3.16, $K = 0.12$ with the same values of ρ. For both values of K, we observe similar phenomena. When ρ is small, our method gives the best result since the area of the positive and the negative values compensate. Although the second order correction of saddle method is better in the conditional case, since all its values remain positive after integration, the overall error is nevertheless larger. When the value of ρ increase. This compensation effect becomes significant in the normal approximation. The correction methods won't improve much of the approximation results, which corresponds to our heuristics.

Figure 3.15: Approximation error of Call function: $n = 100$, $\alpha = 0.05$. $K = 0.03$. $\rho = 0.2$ and $\rho = 0.8$ respectively.

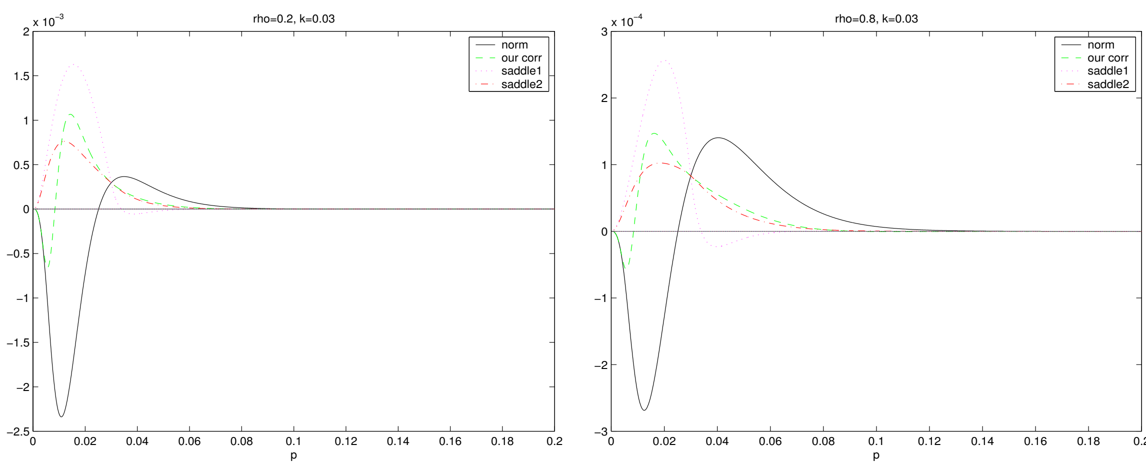
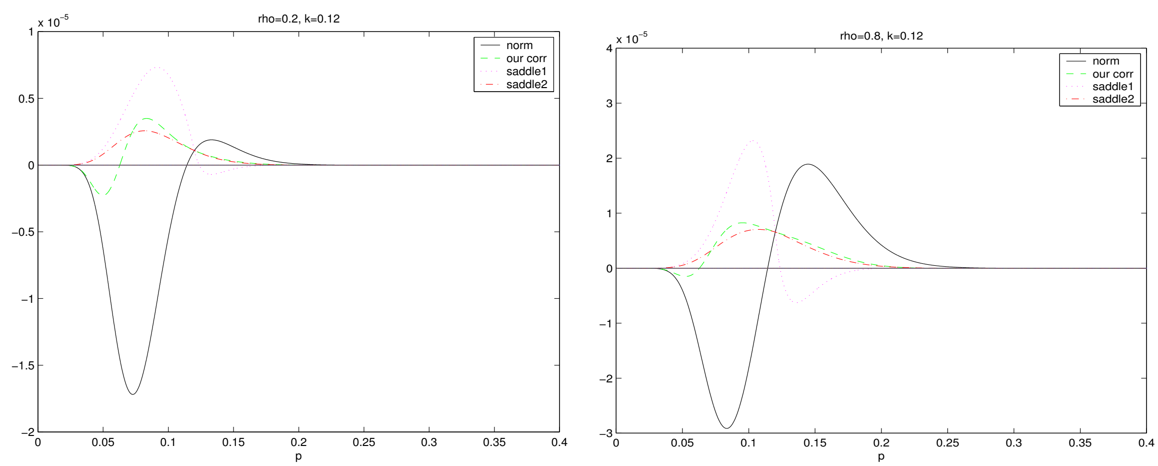


Figure 3.16: $n = 100$, $\alpha = 0.05$. $K = 0.12$. $\rho = 0.2$ and $\rho = 0.8$ respectively.



Chapter 4

Asymptotic Expansions in Normal and Poisson Limit Theorem

This chapter is devoted to a theoretical problem: the asymptotic expansions of $\mathbb{E}[h(W)]$ where W is the sum of independent random variables. It's an extension to Theorem 3.4.8 obtained in the previous chapter. In fact, it is revealed in Chapter 3 that when the function h has certain regularity (for example, f_h has bounded third order derivative), we can improve the normal approximation of $\mathbb{E}[h(W)]$ by proposing a correction term. It is then natural to expect that when h has higher order derivatives with some regularity properties, we may obtain a corrector of the corresponding order. Theoretically speaking, this is a classical problem related to the central limit theorem. By developing some techniques which are based on those in Subsection 3.3.3 in the framework of the Stein's method and the zero bias transformation, we propose an original method to develop the asymptotic expansion. Our main contributions are:

1. We develop a method whose continuous version and discrete version enable respectively to obtain similar results in both the normal and the Poisson context.

2. In the normal case, we give the regularity conditions on h to derive the expansions. For a smooth function, we obtain a full asymptotic expansion. In the general case, the order of the expansion is determined by the derivative order of h. We prove that the approximation error is of appropriate order under some conditions.

3. In the Poisson case, we extend the notion of zero bias transformation to the Poisson distribution. We provide asymptotic expansion for sum of independent random variables and in particular, the independent $0-1$ random variables.

To the finance concern, we improve the approximation of the call function (of course, we can treat other functions) by expanding to higher orders. Numerical results compare the approximation by the second order correction with those obtained by other corrections in the previous chapter and we observe a significant improvement.

One technical difficulty lies in determining the regularity conditions. As we have observed in the previous chapter, we obtain a first order approximation for a function h whose auxiliary function f_h has bounded third order derivative, which means that h itself has up to second order derivatives. However, we have proved that the approximation formula is valid for the call function, which is only absolutely continuous. The approximation error is of the same order. To prove it, we need some refined techniques which involve the usage of the Stein's equation and the concentration inequality. The key point is that although the derivative of the call function, which is an indicator function, is not differentiable, it is nevertheless not so "bad" since it contains only one discontinuous point. In addition to this point, we can obtain the second order derivative of the call function. Therefore, we can apply the standard procedure except at this discontinuous point, which we shall treat by using the concentration inequality.

It is similar in the high order case. In the following, we first provide an expansion formula. We then specify the necessary conditions for the function h when estimating the approximation error. To obtain an N^{th} order expansion, h should have up to N^{th} order derivatives. Moreover, the N^{th} order derivative of h is of finite variation and contains finite number of jump points.

We note that although we obtain a full expansion which merely contains the normal expectations and the Poisson expectations. The formula is presented in recurrence form. So for high order approximations, it requires extra calculations to obtain explicit results. On the other hand, no special effort has been made on the constants of the error estimates.

The chapter is organized as follows: We begin by a brief review (Section 4.1) of the existing results and methods in the literature concerning the asymptotic expansions in the central limit theorem. Section 4.2 and Section 4.3 are devoted respectively to the normal and the Poisson case.

In Subsection 4.2.1, we propose a direct method by taking the Taylor expansion. Since $W^{(i)}$ and X_i, X_i^* are independent, we obtain immediately an expansion formula in recurrence form by replacing the sum variable $W^{(i)}$ with a normal variable. However, since we eliminate one summand variable when considering $W^{(i)}$ instead of W, in the expansion formula, there exist partial sum variables which complicates the calculation. Subsection 4.2.1 contains a refined method of the first one. The key argument is the so called reversed Taylor's formula in the expectation form which allows us to write the terms containing $W^{(i)}$ as some functions of W and X_i, with which we deduce Theorem 4.2.5. The approximation error is estimated in Section 4.2.3. We first propose an estimation procedure and we point out that the growing speed of $\widetilde{f}_h^{(N+2)}$ plays a crucial role for an N^{th} order expansion. The main objective of Section 4.2.3 is to determine under which conditions on h we can deduce desired property of $\widetilde{f}_h^{(N+2)}$. We show that the growing speed of the derivatives is required instead of the boundedness conditions.

In Section 4.3.1, we present the framework of the Stein's method and the zero bias transformation for the Poisson distribution. The similitude in the writing inspired us to study the asymptotic expansion by developing a similar method in the discrete case. The special case of $0-1$ summand variables is first discussed in Section 4.3.2 and the asymptotic expansion is obtained following Lemma 4.3.11. The general case for the sum of independent random variables is shown in Section 4.3.3.

4.1 Introduction

The Stein's method provides a very efficient tool in studying the limit theorems and in developping asymptotic expansions, for both normal and Poisson distributions. For the discrete case of the Poisson distribution, Deheuvels and Pfeifer [21] study the error estimates after one and two ordered estimations for $0-1$ summands by using an operator technique. Borisov and Ruzankin [11] give full asymptotic Poisson expansion for unbounded functions. The usage of Stein's method in the Poisson context is introduced by Chen [15] and is then developed by Barbour and Hall [6], Chen and Choi [19], and Barbour, Chen and Choi [5] for $0-1$ independent random variables. For general nonnegative integer valued summands, Barbour [4] obtains expansions for polynomially growing functions by applying the Stein's method.

In the normal context, the classical method used to derive the normal asymptotic expansions of $\mathbb{E}[h(W)]$ is the Fourier methods as in Hipp [48] and Götze and Hipp [42] by using the Edgeworth's expansion. Barbour [3] used a similar technique in the Poisson case [4] by introducing the Stein's method. This result is extended by Rinott and Rotar [69] and is reviewed in Rotar [71]. In [3], the author considered the expansions for functions with high order derivatives. For a $(l-1)$ times derivable function g, he wrote the expectation $\mathbb{E}[Wg(W)]$ as a sum of l terms containing the cumulant of W and a remaining term, i.e.

$$\mathbb{E}[Wg(W)] = \sum_{k=1}^{l-1} \frac{\xi_{k+1}}{k!}\mathbb{E}[g^{(k)}(W)] + R$$

where ξ_k is the k^{th} order cumulant of W and R is the remainder. The bound of the remaining term contains the cumulant of W and the derivative of g. Combing the equality $\mathbb{E}[h(W) - h(Z)] = \mathbb{E}[Wf_h(W)] - \sigma_W^2\mathbb{E}[f_h'(W)]$, the author obtained by iteration a $(l+1)$-terms expansion of $\mathbb{E}[h(W) - h(Z)]$ and then replaced the term W in the expansion by the normal variable Z. At last, it remains to estimate the error of the above replacement which consists of analytical estimation of the derivatives of f_h with respect to those of h.

From the results of Barbour in [3] and [4], we see that the Stein's method allows us to consider the problem of asymptotic expansions in a similar way for both normal and Poisson approximations. This can also be shown by the results in the following of this chapter. We shall propose a method which adapts to the two cases respectively.

4.2 Normal approximation

4.2.1 The first method

In this subsection, we introduce a first method to get a recurrence form of the asymptotic expansion of $\mathbb{E}[h(W)]$. With the explicit definition of the zero bias transformation, we propose to deal with the expansion of $\mathbb{E}[f_h'(W^*) - f_h'(W)]$ as a whole. The idea is based on the comparison between W and W^* through their common part $W^{(I)}$ and on the independence between $W^{(i)}$, X_i and X_i^*. To be more precise, we consider the Taylor expansion of the above difference at $W^{(i)}$ and then as in [3], replace the sum variables by the normal variables. The expansion formula is of an recurrence form and the proof is proceeded by induction.

We introduce the following notations. Let Θ be the set of indices, i.e. $\Theta = \{1, \cdots, n\}$. For any non-empty subset $J \subset \Theta$, let

$$W_J = \sum_{i \in J} X_i$$

and $\sigma_J^2 = \mathrm{Var}[W_J] = \sum_{i \in J} \sigma_i^2$. In particular, we denote by $W = W_\Theta$ and $\sigma_W = \sigma_\Theta$. Let I_J be a random index of the set J, that is, I_J takes value $i \in J$ with probability $\mathbb{P}(I_J = i) = \sigma_i^2/\sigma_J^2$. We assume in addition that I_J is independent of all X_i and X_i^* for all $i = 1, \cdots, n$. Denote by $J^{(i)} = J \setminus \{i\}$, the subset of J deprived of i. From Proposition 3.2.10, we know that $W_J^{(I_J)} + X_{I_J}^*$ has the W_J-zero biased distribution. In addition, we denote by $f_{h,J} = f_{h,\sigma_J}$ the solution of the Stein's equation $xf(x) - \sigma_J^2 f'(x) = h(x) - \Phi_{\sigma_J}(h)$ and by $f_h = f_{h,\Theta}$.

For any $N \geq 0$, we write $\mathbb{E}[h(W_J)]$ as the sum of two terms: the N^{th}-order estimator $C(J, N, h)$ and the remaining error term $\varepsilon(J, N, h)$. The following theorem gives the recurrence formula to obtain $C(J, N, h)$.

Proposition 4.2.1 *Let $\mathbb{E}[h(W_J)] = C(J, N, h) + e(J, N, h)$. If f_h has up to $(N+2)^{th}$ order derivative, then*

$$C(J,N,h) = C(J,0,h) + \sum_{i \in J} \sigma_i^2 \sum_{k=1}^{N} \frac{1}{k!} C\big(J^{(i)}, N-k, f_{h,J}^{(k+1)}\big) \mathbb{E}\big[(X_i^*)^k - (X_i)^k\big], \quad (4.1)$$

where $C(J,0,h) = \Phi_{\sigma_J}(h)$. *In addition,*

$$|e(J,N,h)| \leq \sum_{i\in J} \sigma_i^2 \Big(\sum_{k=1}^{N} \frac{1}{k!} \big| e\big(J^{(i)}, N-k, f_{h,J}^{(k+1)}\big) \big| \mathbb{E}\big[|(X_i^*)^k - (X_i)^k| \big] + \frac{\|f_{h,J}^{(N+2)}\|_{\sup}}{(N+1)!} \mathbb{E}\big[|X_i^*|^{N+1} + |X_i|^{N+1} \big] \Big). \tag{4.2}$$

Proof. By Stein's equation,

$$h(W_J) - \Phi_{\sigma_J}(h) = W_J f_{h,J}(W_J) - \sigma_J^2 f'_{h,J}(W_J).$$

Taking expectation on the two sides, we have

$$\mathbb{E}[h(W_J)] = \Phi_{\sigma_J}(h) + \sigma_J^2 \mathbb{E}[f'_{h,J}(W_J^*)] - \sigma_J^2 \mathbb{E}[f'_{h,J}(W_J)].$$

By writing the N^{th}-order Taylor expansion of the last two terms at $W_J^{(I_J)}$, the zero-order terms vanish. Moreover, since $W_J^{(i)}$ is independent of X_i and X_i^*, we get

$$\begin{aligned}
\mathbb{E}[h(W_J)] &= \Phi_{\sigma_J}(h) + \sum_{i\in J} \sigma_i^2 \mathbb{E}\big[f'_{h,J}(W_J^{(i)} + X_i^*) - f'_{h,J}(W_J^{(i)} + X_i) \big] \\
&= \Phi_{\sigma_J}(h) + \sum_{i\in J} \sigma_i^2 \sum_{k=1}^{N} \frac{1}{k!} \mathbb{E}\big[f_{h,J}^{(k+1)}(W_J^{(i)}) \big] \mathbb{E}\big[(X_i^*)^k - (X_i)^k \big] + \delta(J,N,h)
\end{aligned} \tag{4.3}$$

where the remaining term

$$\delta(J,N,h) = \sum_{i\in J} \frac{\sigma_i^2}{(N+1)!} \mathbb{E}\big[f_{h,J}^{(N+2)}(W_J^{(i)} + \theta_1 X_i^*)(X_i^*)^{N+1} - f_{h,J}^{(N+2)}(W_J^{(i)} + \theta_2 X_i) X_i^{N+1} \big].$$

Let $C(J,0,h) = \Phi_{\sigma_J}(h)$ and assume we have proved equation (4.1) holds for $l < N$, then we replace $\mathbb{E}\big[f_{h,J}^{(k+1)}(W_J^{(i)}) \big]$ in (4.3) by its $(N-k)^{\text{th}}$-order expansion and get

$$\begin{aligned}
&\mathbb{E}(h(W_J)) \\
&= C(J,0,h) + \sum_{i\in J} \sigma_i^2 \sum_{k=1}^{N} \frac{1}{k!} C\big(J^{(i)}, N-k, f_{h,J}^{(k+1)}\big) \mathbb{E}\big[(X_i^*)^k - (X_i)^k \big] \\
&\qquad + \delta(J,N,h) + \sum_{i\in J} \sigma_i^2 \sum_{k=1}^{N} \frac{1}{k!} e\big(J^{(i)}, N-k, f_{h,J}^{(k+1)}\big) \mathbb{E}\big[(X_i^*)^k - (X_i)^k \big] \\
&= C(J,N,h) + \delta(J,N,h) + \sum_{i\in J} \sigma_i^2 \sum_{k=1}^{N} \frac{1}{k!} e\big(J^{(i)}, N-k, f_{h,J}^{(k+1)}\big) \mathbb{E}\big[(X_i^*)^k - (X_i)^k \big],
\end{aligned}$$

which means that the N^{th}-order estimator is also given by equation (4.1) and the error term is given by

$$e(J,N,h)=\delta(J,N,h)+\sum_{i\in J}\sigma_i^2\sum_{k=1}^{N}\frac{1}{k!}e(J^{(i)},N-k,f_{h,J}^{(k+1)})\mathbb{E}\big[(X_i^*)^k-(X_i)^k\big].$$

Moreover,

$$|\delta(J,N,h)|\leq\sum_{i\in J}\frac{\sigma_i^2}{(N+1)!}\|f_{h,J}^{(N+2)}\|_{\sup}\mathbb{E}\big[|X_i^*|^{N+1}+|X_i|^{N+1}\big].$$

So again by induction, the error term is bounded by

$$\begin{aligned}|e(J,N,h)|\leq&\sum_{i\in J}\sigma_i^2\sum_{k=1}^{N}\frac{1}{k!}\big|e(J^{(i)},N-k,f_{h,J}^{(k+1)})\big|\mathbb{E}\big[\big|(X_i^*)^k-(X_i)^k\big|\big]\\&+\sum_{i\in J}\frac{\sigma_i^2}{(N+1)!}\|f_{h,J}^{(N+2)}\|_{\sup}\Big(\mathbb{E}[|X_i^*|^{N+1}]+\mathbb{E}[|X_i|^{N+1}]\Big).\end{aligned}$$

□

It is apparent that (4.1) is tedious to apply in reality. To do the recurrence for the set J, we need to know the zero order approximation of all its subsets. In addition, we should calculate the normal expectation where the variance is not the same with the variance of the Stein's equation. So Proposition 3.3.24 does not apply here to calculate $\Phi_{\sigma'}(x^m f_{h,\sigma}(x))$, which makes the calculation become much more complicated.

4.2.2 The second method

In this subsection, we propose a refined method to improve the first one in the previous subsection. As shown in the previous subsection, since the Taylor expansion is made around the point $W_J^{(i)}$, at each step, we have to eliminate one variate and calculate a normal expectation function with the reduced variance. This increases significantly the calculation, especially in the exogenous case. Therefore, it's natural to propose a solution by changing in (4.3) the term $\mathbb{E}[f_{h,J}^{(k+1)}(W_J^{(i)})]$ with some expectation function on W_J. This procedure introduces an additional error term. So the objective is to

1) find the relationship between $\mathbb{E}[f(W^{(i)})]$ and $\mathbb{E}[f(W)]$;

2) estimate the error of the above step.

We introduce the following notations. Let X and Y be two independent random variables and let f be a $N+1$ times derivable function. We denote by $\delta(N,f,X,Y)$ the error of the N^{th}-order Taylor's formula in the expectation form, i.e.

$$\mathbb{E}[f(X+Y)]=\sum_{k=0}^{N}\frac{\mathbb{E}[Y^k]}{k!}\mathbb{E}[f^{(k)}(X)]+\delta(N,f,X,Y).\tag{4.4}$$

Recall the Taylor expansion (in [60] for example)

$$f(X+Y)=\sum_{k=0}^{N} f^{(k)}(X)\frac{Y^k}{k!}+\frac{1}{N!}\int_0^1(1-t)^N f^{(N+1)}(X+tY)Y^{N+1}dt. \tag{4.5}$$

By taking the expectation of the above formula, we obtain directly (4.4) since X and Y are independent and we get

$$\delta(N,f,X,Y)=\frac{1}{N!}\int_0^1(1-t)^N\mathbb{E}\big[f^{(N+1)}(X+tY)Y^{N+1}\big]dt, \tag{4.6}$$

or equivalently

$$\delta(N,f,X,Y)=\frac{1}{(N-1)!}\int_0^1(1-t)^{N-1}\mathbb{E}\big[\big(f^{(N)}(X+tY)-f^{(N)}(X)\big)Y^N\big]dt. \tag{4.7}$$

We now present the key formula (4.8) of our method which writes the expectation $\mathbb{E}[f(X)]$ as an expansion of $\mathbb{E}[f(X+Y)]$ and its derivatives multiplied by expectation terms containing the powers of Y. We call (4.8) the *reversed Taylor's formula in the expectation form*. The main feature of this formula is that we treat the products of expectation of functions on random variables $X+Y$ and Y which are not independent. This property makes (4.8) very different with the standard Taylor's formula where (4.4) is obtained by taking expectation of its corresponding form (4.5). However, we show that the remaining term of (4.8) can be deduced from the remaining terms of the standard Taylor's formula.

Proposition 4.2.2 *Let $\varepsilon(N,f,X,Y)$ be the remaining term of the following expansion*

$$\mathbb{E}[f(X)]=\mathbb{E}[f(X+Y)]+\sum_{d\geq 1}(-1)^d\sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}}\mathbb{E}[f^{(|\mathbf{J}|)}(X+Y)]\Big(\prod_{l=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}\Big)+\varepsilon(N,f,X,Y), \tag{4.8}$$

where $|\mathbf{J}|=j_1+\cdots+j_d$ for any $\mathbf{J}=(j_l)\in\mathbb{N}_*^d$. Then, for any integer $N\geq 1$,

$$\varepsilon(N,f,X,Y)=-\sum_{d\geq 0}(-1)^d\sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}}\delta(N-|\mathbf{J}|,f^{(|\mathbf{J}|)},X,Y)\prod_{l=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}. \tag{4.9}$$

Proof. Combining (4.4) and (4.8), we have

$$\begin{aligned}\varepsilon(N,f,X,Y)=&-\sum_{k=1}^{N}\frac{\mathbb{E}[Y^k]}{k!}\mathbb{E}[f^{(k)}(X)]-\delta(N,f,X,Y)\\ &-\sum_{d\geq 1}(-1)^d\sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}}\mathbb{E}[f^{(|\mathbf{J}|)}(X+Y)]\prod_{\lambda=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}.\end{aligned}$$

We take the $(N-|J|)^{\text{th}}$-order Taylor expansion of $\mathbb{E}[f^{(|\mathbf{J}|)}(X+Y)]$ to get

$$\sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\|\mathbf{J}|\leq N}} \mathbb{E}[f^{(|\mathbf{J}|)}(X+Y)]\left(\prod_{\lambda=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}\right)$$

$$= \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\|\mathbf{J}|\leq N}} \left(\mathbb{E}[f^{(|\mathbf{J}|)}(X)] + \sum_{k=1}^{N-|\mathbf{J}|}\frac{\mathbb{E}[Y^k]}{k!}\mathbb{E}[f^{(|\mathbf{J}|+k)}(X)] + \delta(N-|\mathbf{J}|, f^{(|\mathbf{J}|)}, X, Y)\right)\left(\prod_{\lambda=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}\right) \tag{4.10}$$

$$= \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\|\mathbf{J}|\leq N}} \left(\mathbb{E}[f^{(|\mathbf{J}|)}(X)]\left(\prod_{\lambda=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}\right) + \sum_{\substack{\mathbf{J}'=(j_l)\in\mathbb{N}_*^{d+1}\\|\mathbf{J}'|\leq N}}\left(\mathbb{E}[f^{(|\mathbf{J}'|)}(X)]\left(\prod_{\lambda=1}^{d+1}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}\right)\right.\right.$$
$$+ \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\|\mathbf{J}|\leq N}} \delta(N-|\mathbf{J}|, f^{(|\mathbf{J}|)}, X, Y)\left(\prod_{\lambda=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}\right) \tag{4.11}$$

The second term of (4.11) is obtained by regrouping $\frac{\mathbb{E}[Y^k]}{k!}$ in (4.10) with the product term and the sum concerning k with the other sums. Multiplying (4.11) by $(-1)^d$ and taking the sum on d, we notice that most terms disappear and we get

$$\sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\|\mathbf{J}|\leq N}} \mathbb{E}[f^{(|\mathbf{J}|)}(X+Y)]\left(\prod_{\lambda=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}\right)$$
$$= -\sum_{j\leq N}\mathbb{E}[f^{(j)}(X)]\frac{\mathbb{E}[Y^j]}{j!} + \sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\|\mathbf{J}|\leq N}} \delta(N-|\mathbf{J}|, f^{(|\mathbf{J}|)}, X, Y)\left(\prod_{\lambda=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}\right)$$

which implies that

$$\varepsilon(N,f,X,Y) = -\delta(N,f,X,Y) - \sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\|\mathbf{J}|\leq N}} \delta(N-|\mathbf{J}|, f^{(|\mathbf{J}|)}, X, Y)\prod_{l=1}^{d}\frac{\mathbb{E}[Y^{j_l}]}{j_l!}.$$

For simplicity of writing, we write $|J|=0$ when "$J\in\mathbb{N}_*^0$" by convention. In addition, for the empty set $\emptyset$, let $\prod_\emptyset = 1$. With these conventions,

$$\sum_{\substack{\mathbf{J}\in\mathbb{N}_*^0\\|\mathbf{J}|\leq N}} \delta(N-|\mathbf{J}|, f^{(|\mathbf{J}|)}, X, Y)\prod_{l=1}^{0}\frac{\mathbb{E}[Y_{j_l}]}{j_l!} = \delta(N,f,X,Y).$$

Therefore we get (4.9). □

Corollary 4.2.3 *With the notations of (4.4) and (4.8), if f has up to $(N+1)^{th}$ order derivatives and if $f^{(N+1)}$ is bounded, then*

1)

$$|\delta(N, f, X, Y)| \leq \frac{\mathbb{E}[Y^{N+1}]}{(N+1)!} \|f^{(N+1)}\|;$$

2)

$$|\varepsilon(N, f, X, Y)| \leq \|f^{(N+1)}\| \sum_{d \geq 1} \sum_{\substack{\mathbf{J}=(j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}|=N+1}} \prod_{l=1}^{d} \frac{\mathbb{E}[Y^{j_l}]}{j_l!}.$$

Proof. 1) is obvious by definition.
2) From 1) and Proposition 4.2.2, we know that

$$\begin{aligned}
\varepsilon(N, f, X, Y) &\leq \sum_{d \geq 0} \sum_{\substack{\mathbf{J}=(j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \leq N}} |\delta(N - |\mathbf{J}|, f^{(|\mathbf{J}|)}, X, Y)| \prod_{l=1}^{d} \frac{\mathbb{E}[Y^{j_l}]}{j_l!} \\
&\leq \|f^{(N+1)}\| \sum_{d \geq 0} \sum_{\substack{\mathbf{J}=(j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \leq N}} \frac{\mathbb{E}[Y^{N-|J|+1}]}{(N-|J|+1)!} \prod_{l=1}^{d} \frac{\mathbb{E}[Y^{j_l}]}{j_l!}
\end{aligned}$$

which implies 2) by regrouping the product terms. □

Remark 4.2.4
1. Note that δ is relatively easier to study while ε is much more complicated. Therefore, the above proposition facilitates the calculation.

2. The equality (4.8) allows us to write $\mathbb{E}[f(W^{(i)})]$ as an expansion of functions on W. In fact, without specifying the explicit form of ε, one can always propose some expansion form with a remaining term which depends on N, X and Y. The one we propose here is for the purpose to obtain the high order expansion in Theorem 4.2.5.

Before presenting the theorem, we first explain briefly the idea how to replace the terms containing $f(W^{(i)})$ by those of W. Suppose that $\mathbb{E}[f(X)]$ has an expansion

$$\mathbb{E}[f(X)] = \sum_{j=0}^{N} \alpha_j(f, Y) \mathbb{E}[f^{(j)}(X+Y)] + \varepsilon(N, f, X, Y).$$

Then by Taylor's formula, we have

$$
\begin{aligned}
\mathbb{E}[f(X+Y)] &= \mathbb{E}[f(X)] + \sum_{k=1}^{N} \frac{\mathbb{E}[Y^k]}{k!} \mathbb{E}[f^{(k)}(X)] + \delta(N, f, X, Y) \\
&= \mathbb{E}[f(X)] + \delta(N, f, X, Y) \\
&+ \sum_{k=1}^{N} \frac{\mathbb{E}[Y^k]}{k!} \Big(\sum_{j=0}^{N-k} \alpha_j(f^{(k)}, Y) \mathbb{E}[f^{(k+j)}(X+Y)] + \varepsilon(N-k, f^{(k)}, X, Y) \Big).
\end{aligned}
$$

In the last equality, we only need to make the $(N-k)^{\text{th}}$-order expansion of $\mathbb{E}[f^{(k)}(X)]$ to obtain the N^{th}-order expansion of $\mathbb{E}[f(X+Y)]$, . Multiplying by $\mathbb{E}[Y^k]$, we get the sufficient order we need. It follows then

$$
\begin{aligned}
\mathbb{E}[f(X)] = \mathbb{E}[f(X+Y)] &- \sum_{k=1}^{N} \sum_{j=0}^{N-k} \frac{\mathbb{E}[Y^k]}{k!} \alpha_j(f^{(k)}, Y) \mathbb{E}[f^{(k+j)}(X+Y)] \\
&- \sum_{k=1}^{N} \frac{\mathbb{E}[Y^k]}{k!} \varepsilon(N-k, f^{(k)}, X, Y) - \delta(N, f, X, Y).
\end{aligned}
$$

The right-hand side of the above equation consists of an expansion on $X+Y$. The next step is to regroup all the terms of $\mathbb{E}[f^{(l)}(X+Y)]$ for $1 \le l \le N$ to get the expansion. We now present our main theorem.

Theorem 4.2.5 *For any integer $N \ge 0$, we can write $\mathbb{E}[h(W)] = C(N, h) + e(N, h)$ where $C(0, h) = \Phi_{\sigma_W}(h)$ and $e(0, h) = \mathbb{E}[h(W)] - \Phi_{\sigma_W}(h)$, and by induction*

$$
\begin{aligned}
C(N, h) = \Phi_{\sigma_W}(h) + \sum_{i=1}^{n} \sigma_i^2 \sum_{d \ge 1} (-1)^{d-1} &\sum_{\substack{\mathbf{J}=(j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \le N}} C(N - |\mathbf{J}|, f_h^{(|\mathbf{J}|+1)}) \\
&\left(\prod_{l=1}^{d-1} \frac{\mathbb{E}[X_i^{j_l}]}{j_l!} \right) \left(\frac{\mathbb{E}[(X_i^*)^{j_d}]}{j_d!} - \frac{\mathbb{E}[X_i^{j_d}]}{j_d!} \right)
\end{aligned} \tag{4.12}
$$

and for any $N \ge 1$,

$$
\begin{aligned}
& e(N, h) \\
&= \sum_{i=1}^{n} \sigma_i^2 \sum_{d \ge 1} (-1)^{d-1} \sum_{\substack{\mathbf{J}=(j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \le N}} e\big(N - |\mathbf{J}|, f_h^{(|\mathbf{J}|+1)}\big) \left(\prod_{l=1}^{d-1} \frac{\mathbb{E}[X_i^{j_l}]}{j_l!} \right) \left(\frac{\mathbb{E}[(X_i^*)^{j_d}]}{j_d!} - \frac{\mathbb{E}[X_i^{j_d}]}{j_d!} \right) \\
&+ \sum_{i=1}^{n} \sigma_i^2 \sum_{k=0}^{N} \frac{\mathbb{E}[(X_i^*)^k]}{k!} \varepsilon\big(N-k, f_h^{(k+1)}, W^{(i)}, X_i\big) + \sum_{i=1}^{n} \sigma_i^2 \delta\big(N, f_h', W^{(i)}, X_i^*\big)
\end{aligned} \tag{4.13}
$$

if all terms in (4.12) and (4.13) are well defined (here we use the conventions proposed in the proof of Proposition 4.2.2).

Proof. We deduce by induction. The theorem holds when $N = 0$. Suppose that we have proved for $0, \cdots, N-1$ with $N \geq 1$. Recall that $\mathbb{E}[h(W)] = \Phi_{\sigma_W}(h) + \sigma_W^2 \mathbb{E}[f_h'(W^*) - f_h'(W)]$. We shall rewrite $\mathbb{E}[f_h'(W^*)]$ as an expansion on W. By making Taylor expansion of $\mathbb{E}[f_h'(W^{(i)} + X_i^*)]$ at $W^{(i)}$ and then using (4.8) to write the $(N-k)^{\text{th}}$-order expansion of $\mathbb{E}[f_h^{(k+1)}(W^{(i)})]$ as functions of W, we get

$$\begin{aligned}
&\mathbb{E}[f_h'(W^{(i)} + X_i^*)] \\
&= \sum_{k=0}^{N} \frac{\mathbb{E}\big[f_h^{(k+1)}(W^{(i)})\big]}{k!} \mathbb{E}[(X_i^*)^k] + \delta(N, f_h', W^{(i)}, X_i^*) \\
&= \sum_{k=0}^{N} \frac{\mathbb{E}\big[(X_i^*)^k\big]}{k!} \Bigg[\mathbb{E}[f_h^{(k+1)}(W)] + \sum_{d \geq 1} (-1)^d \sum_{\substack{\mathbf{J} = (j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \leq N-k}} \mathbb{E}\big[f_h^{(|\mathbf{J}|+k+1)}(W)\big] \prod_{l=1}^{d} \frac{\mathbb{E}[X_i^{j_l}]}{j_l!} \\
&+ \varepsilon(N-k, f_h^{(k+1)}, W^{(i)}, X_i) \Bigg] + \delta\big(N, f_h', W^{(i)}, X_i^*\big)
\end{aligned} \tag{4.14}$$

Notice that the first term in the bracket when $k = 0$ equals $\mathbb{E}[f_h'(W)]$. For the simplicity of writing, we define the following notation to add the remaining summands when $k \geq 1$ of the first term to the second term as $d = 0$. To be more precise, let by convention

$$\sum_{\substack{\mathbf{J} = (j_l) \in \mathbb{N}_*^0 \\ |\mathbf{J}| \leq N-k}} \mathbb{E}\big[f_h^{(|\mathbf{J}|+k+1)}(W)\big] = \mathbb{E}[f^{(k+1)}(W)]. \tag{4.15}$$

Using the above notation, we can rewrite (4.14) by separating the cases when $k = 0$ and when $k = 1, \cdots, N$ and the remaining terms as

$$\begin{aligned}
&\mathbb{E}[f_h'(W^{(i)} + X_i^*)] - \mathbb{E}[f_h'(W)] \\
&= \sum_{k=1}^{N} \frac{\mathbb{E}[(X_i^*)^k]}{k!} \sum_{d \geq 0} (-1)^d \sum_{\substack{\mathbf{J} = (j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \leq N-k}} \mathbb{E}[f_h^{(|\mathbf{J}|+k+1)}(W)] \prod_{l=1}^{d} \frac{\mathbb{E}[X_i^{j_l}]}{j_l!} &(4.16) \\
&+ \sum_{d \geq 1} (-1)^d \sum_{\substack{\mathbf{J} = (j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \leq N}} \mathbb{E}[f_h^{(|\mathbf{J}|+1)}(W)] \prod_{l=1}^{d} \frac{\mathbb{E}[X_i^{j_l}]}{j_l!} &(4.17) \\
&+ \sum_{k=0}^{N} \frac{\mathbb{E}[(X_i^*)^k]}{k!} \varepsilon(N-k, f_h^{(k+1)}, W^{(i)}, X_i) + \delta(N, f_h', W^{(i)}, X_i^*). &(4.18)
\end{aligned}$$

By interchanging summations, we have

$$(4.16) = \sum_{d \geq 0} (-1)^d \sum_{k=1}^{N} \sum_{\substack{\mathbf{J} = (j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \leq N-k}} \mathbb{E}[f_h^{(|\mathbf{J}|+k+1)}(W)] \frac{\mathbb{E}[(X_i^*)^k]}{k!} \prod_{l=1}^{d} \frac{\mathbb{E}[X_i^{j_l}]}{j_l!}.$$

We then regroup $\frac{\mathbb{E}[(X_i^*)^k]}{k!}$ with the product term to get

$$(4.16) = \sum_{d\geq 0}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^{d+1}\\ |\mathbf{J}|\leq N}} \mathbb{E}[f_h^{(|\mathbf{J}|+1)}(W)]\frac{\mathbb{E}[(X_i^*)^{j_{d+1}}]}{(j_{d+1})!}\prod_{l=1}^{d}\frac{\mathbb{E}[X_i^{j_l}]}{j_l!}$$

$$= \sum_{d\geq 1}(-1)^{d-1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^{d}\\ |\mathbf{J}|\leq N}} \mathbb{E}[f_h^{(|\mathbf{J}|+1)}(W)]\frac{\mathbb{E}[(X_i^*)^{j_d}]}{(j_d)!}\prod_{l=1}^{d-1}\frac{\mathbb{E}[X_i^{j_l}]}{j_l!}$$

Therefore, taking the sum of (4.16) and (4.17), we get

$$\begin{aligned}&(4.16)+(4.17)\\ &= \sum_{d\geq 1}(-1)^{d-1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^{d}\\ |\mathbf{J}|\leq N}} \mathbb{E}[f_h^{(|\mathbf{J}|+1)}(W)]\left(\prod_{l=1}^{d-1}\frac{\mathbb{E}[X_i^{j_l}]}{j_l!}\right)\left(\frac{\mathbb{E}[(X_i^*)^{j_d}]}{(j_d)!}-\frac{\mathbb{E}[(X_i)^{j_d}]}{(j_d)!}\right)\end{aligned}$$

At last, since we have proved the theorem for all $N-|\mathbf{J}|<N$, we replace $\mathbb{E}[f_h^{(|\mathbf{J}|+1)}(W)]$ by its $(N-|\mathbf{J}|)^{\text{th}}$ order expansion $C(N-|\mathbf{J}|, f_h^{(|\mathbf{J}|+1)}) + e(N-|\mathbf{J}|, f_h^{(|\mathbf{J}|+1)})$ to obtain (4.12). Finally, it suffices to notice that $e(N,h)$ contains the terms in (4.18) and the terms in the above replacement of lower orders. □

Corollary 4.2.6 *The expansion of the first two orders are given by*

1)

$$C(1,h) = \Phi_{\sigma_W}(h) + \Phi\Big(\big(\frac{x^3}{3\sigma_W^4}-\frac{x}{\sigma_W^2}\big)h(x)\Big)\mathbb{E}[X_I^*]$$

2)

$$\begin{aligned}C(2,h) = C(1,h) &+ \Phi_{\sigma_W}\left(\Big(\frac{x^6}{18\sigma_W^8}-\frac{5x^4}{6\sigma_W^6}+\frac{5x^2}{2\sigma_W^4}\Big)(h(x)-\Phi_{\sigma_W}(h))\right)\mathbb{E}[X_I^*]^2\\ &+\frac{1}{2}\Phi_{\sigma_W}\left(\Big(\frac{x^4}{4\sigma_W^6}-\frac{3x^2}{2\sigma_W^4}\Big)h(x)\right)\big(\mathbb{E}[(X_I^*)^2]-\mathbb{E}[X_I^2]\big).\end{aligned} \tag{4.19}$$

Proof. 1) is a direct result of the above proposition.
2) By (4.12),

$$C(2,h) = \Phi_{\sigma_W}(h) + \sum_{i=1}^n \sigma_i^2\Big(C(1,f_h'')\mathbb{E}[X_i^*] + C(0,f_h^{(3)})\frac{\mathbb{E}[(X_i^*)^2-X_i^2]}{2}\Big).$$

Then it suffices to calculate $\Phi_{\sigma_W}\big((\frac{x^3}{3\sigma_W^4}-\frac{x}{\sigma_W^2})f_h''\big)$ and $\Phi_{\sigma_W}(f_h^{(3)})$ by Proposition 3.3.24.
□

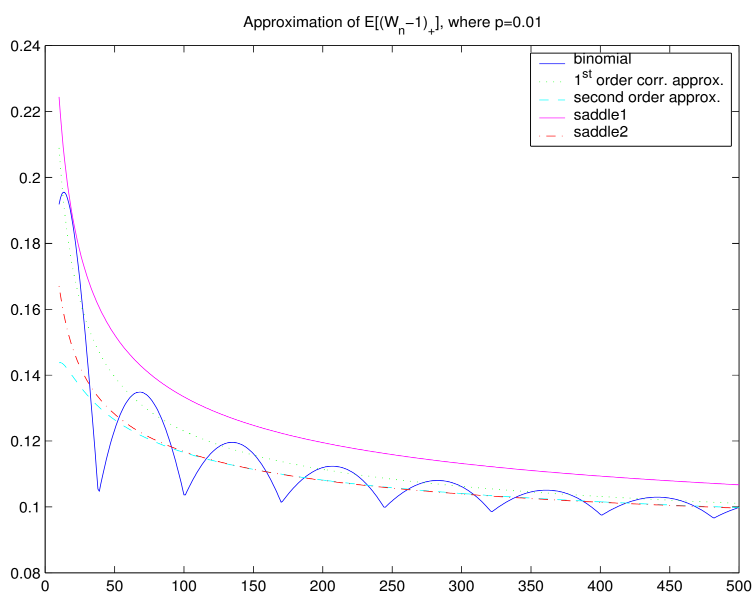
Remark 4.2.7 The first order correction given by Theorem 3.4.8 is a special case here when $N = 1$.

4.2.2.1 Numerical result

We apply (4.19) to the call function and we present the numerical results for i.i.d. random variables X_i. Figure 4.1 and 4.2 compare the second order approximation $C(2,h)$ to other approximations: the first order approximation $\Phi_{\sigma_W}(h) + C(1,h)$, the first and the second order approximations by the saddle point method. The test is the same with that for Figure 3.10 and 3.11. We observe that $C(2,h)$ provides better approximation quality than $C(1,h)$. It is of the same precision of the second order saddle-point approximation.

In this case, similar with the first order approximation of the indicator function. We can not obtain theoretically the approximation error estimation since the call function is only one time differentiable. The explanation of this improvement should be similar to that for the indicator function case.

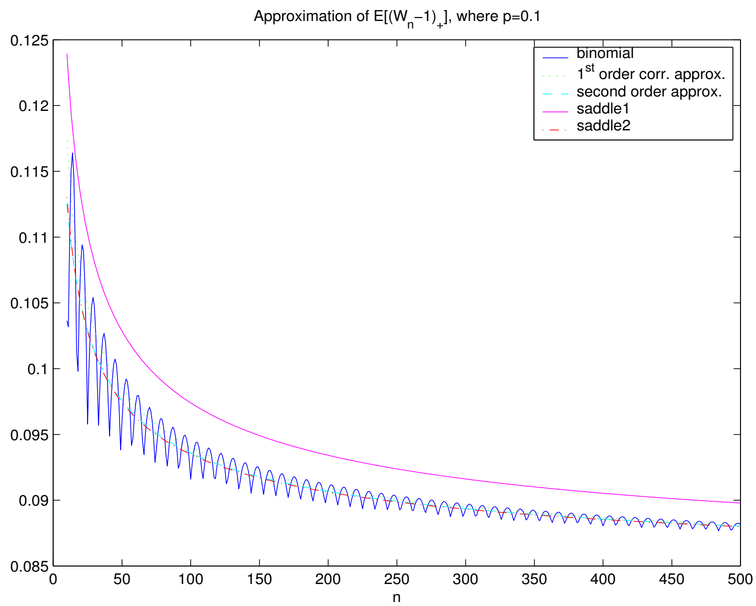
Figure 4.1: Second order expansion for Call function, asymptotic case: $p = 0.01$, and $k = 1$.



4.2.3 Estimation of the approximation error

In this subsection, we estimate the error bound $e(N,h)$ given by (4.13). The idea is as below.

Figure 4.2: Second order expansion for Call function: $p = 0.1$ and $k = 1$. The second order approximation of our method and the saddle-point method coincide.



Procedure to estimate $e(N, h)$:

1) The error term $e(N, h)$ contains two types of terms. The first one consists of the errors of lower orders $e(N-k, f_h^{(k+1)})$ where $k = 1, \cdots, N$. These terms can be bounded by induction once the estimation has been established for $1, \cdots, N-1$. The other terms to estimate are $\delta(N, f_h', W^{(i)}, X_i^*)$ and $\varepsilon(N-k, f_h^{(k+1)}, W^{(i)}, X_i)$. By Proposition 4.2.2,

$$\sum_{k=0}^{N} \frac{\mathbb{E}[(X_i^*)^k]}{k!} \varepsilon\big(N-k, f_h^{(k+1)}, W^{(i)}, X_i\big)$$

$$= -\sum_{k=0}^{N} \frac{\mathbb{E}[(X_i^*)^k]}{k!} \sum_{d\geq 0} (-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} \delta(N-k-|\mathbf{J}|, f_h^{(k+|\mathbf{J}|+1)}, W^{(i)}, X_i) \prod_{l=1}^{d} \frac{\mathbb{E}[X_i^{j_l}]}{j_l!}$$

$$= \sum_{d\geq 1} (-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} \delta(N-|\mathbf{J}|, f_h^{(|\mathbf{J}|+1)}, W^{(i)}, X_i) \Big(\prod_{l=1}^{d-1} \frac{\mathbb{E}[X_i^{j_l}]}{j_l!} \Big) \frac{\mathbb{E}[(X_i^*)^{j_d}]}{j_d!}$$

Then, it suffices to consider the term $\delta(N-k, f_h^{(k+1)}, X, Y)$ where $k = 0, 1, \cdots, N$ for independent random variables X and Y to estimate the second type of terms.

2) By the explicit formula (4.6) of $\delta(N,f,X,Y)$, we have

$$\delta(N-k,f_h^{(k+1)},X,Y)=\frac{1}{(N-k)!}\int_0^1(1-t)^{N-k}\mathbb{E}[f_h^{(N+2)}(X+tY)Y^{N-k+1}]dt.$$

If the growing speed of $f_h^{(N+2)}$ is controlled, then $\delta(N-k,f_h^{(k+1)},X,Y)$ can be bounded by some quantity containing the moments of X and Y. To be more precise, if $|f_h^{(N+2)}|\leq c|x|^m+d$ where c and d are some constants and m is some integer, then

$$\begin{aligned}
&|\delta(N-k,f_h^{(k+1)},X,Y)|\\
&\leq\frac{1}{(N-k)!}\int_0^1(1-t)^{N-k}\mathbb{E}\big[(c|X+tY|^m+d)|Y|^{N-k+1}\big]dt\\
&\leq\frac{c}{(N-k)!}\sum_{l=0}^m\binom{m}{l}\mathbb{E}[|X|^l]\mathbb{E}[|Y|^{N+m-l-k+1}]\int_0^1(1-t)^{N-k}t^{m-l}dt\\
&+\frac{d}{(N-k)!}\mathbb{E}[|Y|^{N-k+1}]\int_0^1(1-t)^{N-k}dt\\
&=\sum_{l=0}^m\frac{m!c}{l!(N+m-l-k+1)!}\mathbb{E}[|X|^l]\mathbb{E}[|Y|^{N+m-l-k+1}]+\frac{d}{(N-k+1)!}\mathbb{E}[|Y|^{N-k+1}].
\end{aligned}$$

The last equality is because

$$\int_0^1(1-t)^{N-k}t^{m-l}dt=\frac{(m-l)!(N-k)!}{(N+m-l-k+1)!}.$$

We replace X by $W^{(i)}$ and Y by X_i or X_i^*. The leading order concerning the moments of $|X_i|$ or $|X_i^*|$ is $N-k+1$. Hence if we suppose, in addition, that W has finite moments up to order m, the estimation is of the right order, that is, the approximation error of the N^{th} order expansion is of the same order with $\mathbb{E}[|X_i|^{N+1}]$ and $\mathbb{E}[|X_i^*|^{N+1}]$, which is $O\big(\frac{1}{\sqrt{n}^{N+1}}\big)$ in the binomial case. Note that the growing speed m of $|f_h^{(N+2)}|$ intervenes in the moment condition of the sum variable W.

3) Estimate the growing speed of $|f_h^{(N+2)}|$ with respect to that of h.

4) Discuss the first type of terms $e(N-k,f_h^{(k+1)})$ using estimations in the above steps.

Therefore, our main objective is to estimate $|f_h^{(N+2)}|$. We shall develop techniques which have been introduced in the subsection 3.3.3 and we work with the function $\widetilde{f}_h$ defined by (3.28) instead of with f_h directly since the derivative functions are not necessarily centralized.

In the following, we first present the necessary conditions for h under which we can deduce the N^{th}-order expansion. We then study respectively the regularity of $\widetilde{f}_h$ given the function h and the growing speed of $\widetilde{f}_h^{(N+2)}$ at infinity. At last, we summarize these result to discuss the estimations of $e(N-k,f_h^{(k+1)})$ at the end of this subsection.

4.2.3.1 Conditions on h for the N^{th}-order expansion

The objective here is to determine, given an integer $N > 0$, for which functions h we can take the N^{th}-order expansion.

In subsection 3.3.2, we have introduced the vector space $\mathscr{E}$ for $\widetilde{f}_h$. The definition of this function set specifies in fact the functions we are interested in. We now extend the notion to a larger context. Let $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$ be the set of all non-zero real numbers. Let $\mathscr{H}_\sigma^0$ be the vector space of all functions h defined on $\mathbb{R}_*$ which are locally of finite variation and have finite number of jumps such that

$$\int \big|P(x)h(x)\big|\phi_\sigma(x)1\!\!1_{\{|x|>a\}}dx < +\infty$$

for any polynomial $P(x)$ and any real number $a > 0$. Notice that $\mathscr{H}_\sigma^0$ is a subset of $\mathscr{E}$ and contains all functions in $\mathscr{E}$ which are of polynomial increasing speed at infinity. Obviously, if $h \in \mathscr{H}_\sigma^0$, then $\widetilde{f}_h$ is well defined. Compared to the definition of $\mathscr{E}$, the additional condition concerning $P(x)$ is for purpose that the increasing speed of the auxiliary function $\widetilde{f}_h$ is controlled.

In Theorem 3.4.8, we have supposed the boundedness condition of $f_h^{(3)}$ to estimate the first order approximation error. In fact, this condition can be relaxed to the "call function" whose $f_h^{(3)}$ does not exist. This result leads us to propose, in the general case, for any integer $N > 0$,

$$\mathscr{H}_\sigma^N = \{h \mid h : \mathbb{R}_* \to \mathbb{R} \text{ having up to } N^{\text{th}} \text{ order derivatives such that } h^{(N)} \in \mathscr{H}_\sigma^0\}.$$

The indicator function I_α belongs to $\mathscr{H}_\sigma^0$ and the call function C_k belongs to $\mathscr{H}_\sigma^1$. Heuristically, we shall make N^{th}-order expansion for functions in the set $\mathscr{H}_\sigma^N$.

4.2.3.2 The regularity of $\widetilde{f}_h$

In this subsection, we shall prove that $\widetilde{f}_h \in \mathscr{H}_\sigma^{N+1}$ if $h \in \mathscr{H}_\sigma^N$. Then, by definition of $\mathscr{H}_\sigma^0$, we know that $\widetilde{f}_h^{(N+1)}$ has finite number of jumps. Therefore, when $|x|$ is large enough, $\widetilde{f}_h^{(N+2)}$ is well defined.

To prove this, the idea is more or less shown in the proof of Proposition 3.3.20. We now summarize and generalize it as follows.

Procedure to estimate $\widetilde{f}_h^{(N+1)}$*:*

i) we generalize Corollary 3.3.18 to write $\widetilde{f}_h^{(N)}$ as a function of $\widetilde{f}_{\Gamma(h)}, \widetilde{f}_{\Gamma^2(h)}, \cdots, \widetilde{f}_{\Gamma^N(h)}$, (Proposition 4.2.10);

ii) we write $\Gamma^N(h)$ as a function of $h, h', \cdots, h^{(N)}$, (Proposition 4.2.11);

iii) we study $\widetilde{f}_h^{(N+1)}$ using i) and ii), (Proposition 4.2.13).

We remark that the above procedure proposes a method which provides the suitable order of the estimation. No special effort is made to improve the constants in the upper bound in this chapter.

We first show that $\widetilde{f}_{\Gamma^N(h)}$ is well defined if $h \in \mathscr{H}_\sigma^N$.

Proposition 4.2.8 *For any integer $N \geq 1$, we have $\mathscr{H}_\sigma^N \subset \mathscr{H}_\sigma^{N-1}$. If $h \in \mathscr{H}_\sigma^N$, then*

1) $\widetilde{f}_h$ is well defined;

2) $\Gamma(h) \in \mathscr{H}_\sigma^{N-1}$;

3) $\widetilde{f}_{\Gamma^N(h)}$ is well defined.

Proof. If $h \in \mathscr{H}_\sigma^N$, then $h^{(N-1)}$ is a continuous function. Let $a > 0$ be a real number. Then $h^{(N-1)}(x) = h^{(N-1)}(a) + \displaystyle\int_a^x h^{(N)}(t)dt$ for $x \geq a$. Let $P(x)$ be any polynomial. By Fubini's theorem

$$\begin{aligned}
&\int_a^\infty |h^{(N-1)}(x)P(x)|\phi_\sigma(x)dx \\
\leq\; &|h^{(N-1)}(a)| \int_a^{+\infty} |P(x)|\phi_\sigma(x)dx + \int_a^\infty dt|h^{(N)}(t)| \int_t^{+\infty} |P(x)|\phi_\sigma(x)dx.
\end{aligned}$$

By integration by part, there exists another polynomial $Q(t)$ such that

$$\int_t^{+\infty} |P(x)|\phi_\sigma(x)dx \leq Q(t)\phi_\sigma(t)$$

for any $t \geq a$. So we know that $\int_a^\infty |h^{(N-1)}(x)P(x)|\phi_\sigma(x)dx < +\infty$. Similarly we can prove that $\displaystyle\int_{-\infty}^{-a} |P(x)h^{(N-1)}(x)|\phi_\sigma(x)dx < +\infty$. Therefore $f \in \mathscr{H}_\sigma^{N-1}$.

1) If $h \in \mathscr{H}_\sigma^N$, then by the argument above, $h \in \mathscr{H}_\sigma^0$. So $\widetilde{f}_h$ is well defined.

2) By definition, if h has up to N^{th} order derivatives, then $\Gamma(h)$ has up to $(N-1)^{\text{th}}$ derivatives and

$$\big(\Gamma(h)\big)^{(N-1)}(x) = \left(\frac{h(x)}{x}\right)^{(N)} = \sum_{k=0}^N (-1)^{N-k} \frac{N!}{k!} \frac{h^{(k)}(x)}{x^{N-k+1}}.$$

Then for any polynomial $P(x)$,

$$\int \big|\Gamma(h)^{(N-1)}(x)P(x)\big|\phi_\sigma(x)1\!\!1_{\{|x|>a\}}dx \leq \sum_{k=0}^N \frac{N!}{k!a^{N-k+1}} \int \big|P(x)h^{(N-k)}(x)\big|\phi_\sigma(x)1\!\!1_{\{|x|>a\}}dx.$$

Since $h \in \mathscr{H}_\sigma^N \subset \mathscr{H}_\sigma^{N-1} \subset \cdots \subset \mathscr{H}_\sigma^0$, the integral above is finite, therefore $\Gamma(h) \in \mathscr{H}_\sigma^{N-1}$.

3) By 2), $\Gamma^N(h) \in \mathscr{H}_\sigma^0$, which deduces 3). $\square$

Remark 4.2.9 From the proof of Proposition 4.2.8 we observe that if f is a function in $\mathscr{H}_\sigma^N$, then for any polynomial $P(x)$, $P(x)f(x) \in \mathscr{H}_\sigma^N$. Furthermore $f(x)/x \in \mathscr{H}_\sigma^N$ and $f'(x) \in \mathscr{H}_\sigma^{N-1}$.

We have shown in Corollary 3.3.18 that $\widetilde{f}_h'(x) = x\widetilde{f}_{\Gamma(h)}(x)$ and we've remarked that since $\Gamma(h)$ grows more slowly at infinity than h, this equality enables us to get the estimations of the right order which are difficult to obtain by using directly the Stein's equation. On the other hand, the above equality yields that

$$\widetilde{f}_h''(x) = \big(x\widetilde{f}_{\Gamma(h)}(x)\big)' = x^2\widetilde{f}_{\Gamma^2(h)}(x) + \widetilde{f}_{\Gamma(h)}(x);$$

and

$$\widetilde{f}_h^{(3)}(x) = x^3\widetilde{f}_{\Gamma^3(h)} + 3x\widetilde{f}_{\Gamma^2(h)}$$

and so on, which suggests the existence of a general formula of $\widetilde{f}_h^{(N)}$. The following proposition gives this result.

Proposition 4.2.10 *For any $h \in \mathscr{H}_\sigma^N$ we have the following equality:*

$$\widetilde{f}_h^{(N)}(x) = \sum_{k=0}^{[N/2]} \binom{N}{2k}(2k-1)!!x^{N-2k}\widetilde{f}_{\Gamma^{N-k}(h)}(x). \tag{4.20}$$

Here we use the convention $(-1)!! = 1$

Proof. The equality (4.20) is clearly true when $N = 0$. Suppose that we have verified (4.20) for $0, \cdots, N$, then for $N+1$, we have

$$\begin{aligned}
\widetilde{f}_h^{(N+1)}(x) = (\widetilde{f}_h^{(N)}(x))' &= \sum_{k\geq 0}\binom{N}{2k}(2k-1)!!(x^{N-2k}\widetilde{f}_{\Gamma^{N-k}(h)})' \\
= \sum_{k\geq 0}\binom{N}{2k}(2k-1)!!x^{N-2k+1}\widetilde{f}_{\Gamma^{N-k+1}(h)}(x) & \qquad (4.21)\\
+\sum_{k\geq 0}\binom{N}{2k}(2k-1)!!\Big((N-2k)x^{N-2k-1}\widetilde{f}_{\Gamma^{N-k}(h)}\Big) & \qquad (4.22)
\end{aligned}$$

Changing the index $l = k+1$ in (4.22), we get

$$(4.22) = \sum_{l\geq 1}\binom{N}{2l-2}(2l-3)!!\Big((N-2l+2)x^{N-2l+1}\widetilde{f}_{\Gamma^{N-l+1}(h)}\Big).$$

Hence

$$\begin{aligned}
&(4.21)+(4.22) = x^{N+1}\widetilde{f}_{\Gamma^{N+1}(h)}(x)\\
&+\sum_{l\geq 1}\left[\binom{N}{2l-2}(2l-3)!!(N-2l+2)+\binom{N}{2l}(2l-1)!!\right]x^{N-2l+1}\widetilde{f}_{\Gamma^{N-l+1}(h)}(x),
\end{aligned}$$

On the other hand,

$$
\begin{aligned}
&\binom{N}{2l-2}(2l-3)!!(N-2l+2)+\binom{N}{2l}(2l-1)!! \\
=&\frac{N!}{(2l-2)!(N-2l)!}(2l-3)!\Big(\frac{1}{N-2l+1}+\frac{1}{2l}\Big) \\
=&\frac{N!}{(2l-1)!(N-2l)!}(2l-1)!!\frac{N+1}{2l(N-2l+1)}=\binom{N+1}{2l}(2l-1)!!.
\end{aligned}
$$

As a consequence,

$$
\begin{aligned}
\widetilde{f}_h^{(N+1)}(x)&=x^{N+1}\widetilde{f}_{\Gamma^{N+1}(h)}(x)+\sum_{l\geq 1}\binom{N+1}{2l}(2l-1)!!x^{N+1-2l}\widetilde{f}_{\Gamma^{N+1-l}(h)}(x) \\
&=\sum_{l\geq 0}\binom{N+1}{2l}(2l-1)!!x^{N+1-2l}\widetilde{f}_{\Gamma^{N+1-l}(h)}(x),
\end{aligned}
$$

which ends the proof by induction. $\square$

We are now interested in the growing speed of $\Gamma^N(h)$ with respect to the derivative $h^{(N)}$. By definition,

$$
\Gamma(h)=\Big(\frac{h(x)}{x}\Big)'=\frac{h'(x)}{x}-\frac{h(x)}{x^2}.
$$

Then

$$
\Gamma^2(h)=\bigg(\frac{\big(\Gamma(h)(x)\big)}{x}\bigg)'=\frac{h''(x)}{x^2}-\frac{3h'(x)}{x^3}+\frac{3h(x)}{x^4}.
$$

The general formula is given below.

Proposition 4.2.11 *For any $h\in\mathscr{H}_\sigma^N$,*

$$
\Gamma^N(h)(x)=\sum_{k=0}^{N}(-1)^k(2k-1)!!\binom{N+k}{2k}\frac{h^{(N-k)}(x)}{x^{N+k}}. \tag{4.23}
$$

By convention, $(-1)!!=1$.

Proof. We will prove the theorem by induction. When $N=0$, the theorem is evident. Suppose that we have proved for $0,\cdots,N$, then by the linearity of Γ, we have

$$
\begin{aligned}
\Gamma^{N+1}(h)=\Gamma(\Gamma^N(h))&=\sum_{k=0}^{N}(-1)^k(2k-1)!!\binom{N+k}{2k}\Gamma\Big(\frac{h^{(N-k)}}{x^{N+k}}\Big) \\
&=\sum_{k=0}^{N}(-1)^k(2k-1)!!\binom{N+k}{2k}\Big(\frac{h^{(N-k)}(x)}{x^{N+k+1}}\Big)'.
\end{aligned}
$$

We write $\Gamma^{N+1}(h)$ as the sum of two terms

$$\Gamma^{N+1}(h) = \sum_{k=0}^{N} (-1)^k (2k-1)!! \binom{N+k}{2k} \frac{h^{(N-k+1)}(x)}{x^{N+k+1}} \tag{4.24}$$

$$+ \sum_{k=0}^{N} (-1)^{k+1} (2k-1)!! \binom{N+k}{2k} (N+k+1) \frac{h^{(N-k)}(x)}{x^{N+k+2}}. \tag{4.25}$$

Changing the index $l = k+1$ in (4.25), we obtain

$$(4.25) = \sum_{l=1}^{N+1} (-1)^l (2l-3)!! \binom{N+l-1}{2l-2} (N+l) \frac{h^{(N-l+1)}(x)}{x^{N+l+1}}.$$

Then

$$\Gamma^{N+1}(h) = \frac{h^{(N+1)}(x)}{x^{N+1}} + (-1)^{N+1}(2N+1)!! \frac{h(x)}{x^{2N+2}}$$
$$+ \sum_{l=1}^{N} (-1)^l \left[(2l-1)!! \binom{N+l}{2l} + (2l-3)!! \binom{N+l-1}{2l-2} (N+l) \right] \frac{h^{(N-l+1)}(x)}{x^{N+l+1}}.$$

Notice that

$$\begin{aligned} &(2l-1)!! \binom{N+l}{2l} + (2l-3)!! \binom{N+l-1}{2l-2} (N+l) \\ &= (2l-3)!! \frac{(N+l)!}{(2l)!(N-l+1)!} \Big((2l-1)(N-l+1) + 2l(2l-1) \Big) \\ &= (2l-1)!! \binom{N+l+1}{2l}. \end{aligned}$$

Then we get

$$\begin{aligned} \Gamma^{N+1}(h) &= \frac{h^{(N+1)}(x)}{x^{N+1}} + \sum_{l=1}^{N+1} (-1)^l (2l-1)!! \binom{N+1+l}{2l} \frac{h^{(N-l+1)}(x)}{x^{N+l+1}} \\ &= \sum_{l=0}^{N+1} (-1)^l (2l-1)!! \binom{N+1+l}{2l} \frac{h^{(N+1-l)}(x)}{x^{N+1+l}}. \end{aligned}$$

By induction we have proved the theorem. □

Our objective is to estimate $\widetilde{f}_h^{(N+2)}$ for a function $h \in \mathscr{H}_\sigma^N$. From Proposition 4.2.10 and Proposition 4.2.11, we can write $\widetilde{f}_h^{(N)}$ as some linear combination of the terms $\widetilde{f}_{\frac{h^{(N-l)}}{x^{N+l}}}$ where $l = 0, 1, \cdots, N$. However, we can not estimate these result directly since we know the properties on $h^{(N+1)}$, but not on $h^{(N+2)}$ and there will be one term we can not study. The solution to this problem is similar with that for the "call" function, that is, we calculate $\widetilde{f}_h^{(N+1)}$ by using Proposition 4.2.10 and Proposition 4.2.11. Then we use the Stein's equation to get $\widetilde{f}_h^{(N+2)}$. This estimation is given by Proposition 4.2.17.

Lemma 4.2.12 *If $h \in \mathscr{H}_\sigma^0$, then $\widetilde{f}_h \in \mathscr{H}_\sigma^1$.*

Proof. By Stein's equation, $\widetilde{f}_h' = \frac{1}{\sigma^2}(x\widetilde{f}_h - h)$. Since $h \in \mathscr{H}_\sigma^0$, we know that $\widetilde{f}_h'$ is locally of finite variation and has finite number of jumps. It suffices to verify that $\int |P(x)x\widetilde{f}_h(x)|\phi_\sigma(x)|1\!\!1_{\{|x|>a\}}dx < \infty$. In fact, by (3.28), for any $a > 0$,

$$\begin{aligned}\int_a^\infty |P(x)x\widetilde{f}_h(x)|\phi_\sigma(x)dx &\leq \frac{1}{\sigma^2}\int_a^\infty |P(x)x|\Big(\int_x^\infty |h(t)|\phi_\sigma(t)dt\Big)dx\\ &\leq \frac{1}{\sigma^2}\int_a^\infty \Big(\int_a^t |P(x)x|dx\Big)|h(t)|\phi_\sigma(t)dt.\end{aligned}$$

We know that there exists some polynomial function Q such that $\int_a^t |P(x)x|dx \leq Q(t)$. Since $h \in \mathscr{H}_\sigma^0$, we know that the above integral is finite. The case when $a < 0$ is similar, which follows the lemma. □

Proposition 4.2.13 *For any $h \in \mathscr{H}_\sigma^N$, $\widetilde{f}_h \in \mathscr{H}_\sigma^{N+1}$.*

Proof. We need to prove that $\widetilde{f}_h^{(N+1)} \in \mathscr{H}_\sigma^0$. By using Proposition 4.2.10 and separating the first term with the others, we have

$$\widetilde{f}_h^{(N)}(x) = x^N\widetilde{f}_{\Gamma^N(h)}(x) + \sum_{k=1}^{[N/2]}\binom{N}{2k}(2k-1)!!x^{N-2k}\widetilde{f}_{\Gamma^{N-k}(h)}(x).$$

Since $\Gamma^N(h) \in \mathscr{H}_\sigma^0$, we only need to discuss the derivative of the first term. By Lemma 4.2.12, $\widetilde{f}_{\Gamma^N(h)} \in \mathscr{H}_\sigma^1$. Therefore, $\widetilde{f}_h^{(N)} \in \mathscr{H}_\sigma^1$, which implies that $\widetilde{f}_h^{(N+1)}$ belongs to $\mathscr{H}_\sigma^0$. □

Remark 4.2.14 Suppose that h is a function in $\mathscr{H}_\sigma^0$ which agrees with a function in $\mathscr{H}_\sigma^1$ when $|x|$ is sufficiently large. (It is equivalent to suppose that the continuous part is in $\mathscr{H}_\sigma^1$). Then $\widetilde{f}_h'$ agrees with a function in $\mathscr{H}_\sigma^1$ when $|x|$ is sufficiently large. In fact, by Stein's equation, $\widetilde{f}_h' = \frac{1}{\sigma^2}(xf_h(x) - h(x))$. Since $xf_h(x) \in \mathscr{H}_\sigma^1$, we know that f_h' agrees with a function in $\mathscr{H}_\sigma^1$ when $|x|$ sufficiently large since it is the case for h. More generally, if h is a function in $\mathscr{H}_\sigma^N$ such that $h^{(N)}$ agrees with a function in $\mathscr{H}_\sigma^1$ (i.e. h agrees with a function in $\mathscr{H}_\sigma^{N+1}$) when $|x|$ is sufficiently large, then it is the same thing for $\widetilde{f}_h^{(N+1)}$. Therefore, we can study the behavior of $\widetilde{f}_h^{(N+2)}$ at infinity.

4.2.3.3 The growing speed of $\widetilde{f}_h^{(N+2)}$ at infinity

In the following we shall study the increasing speed (the order) of the auxiliary function $\widetilde{f}_h^{(N+2)}$ when knowing the order of h and its derivatives. Here we are only interested in the order of the error estimation, but not the bound constant.

Lemma 4.2.15 *Let $h \in \mathscr{H}_\sigma^N$ such that $h^{(N)}(x) = O(|x|^m)$ $(x \to \infty)$, where $m \geq 0$ is an integer. Then for any integer $0 \leq k \leq N$,*

1) $h^{(N-k)}(x) = O(|x|^{m+k})$;

2) $\Gamma^{N-k}(h)(x) = O(|x|^{m-N+2k})$

Proof. 1) By induction it suffices to prove that $h^{(N-1)}(x) = O(|x|^{m+1})$. In fact, for any $x > 1$, we have

$$h^{(N-1)}(x) = h^{(N-1)}(1) + \int_1^x h^{(N)}(t)dt = O(x^{m+1}).$$

The case when $x < -1$ is similar.
2) By Proposition 4.2.11 and 1) we know that

$$\Gamma^{N-k}(h) = \sum_{j=0}^{N-k} O\Big(\frac{|x|^{m+k+j}}{|x|^{N-k+j}}\Big) = O(|x|^{m-N+2k}).$$

□

Lemma 4.2.16 *Let $h \in \mathscr{H}_\sigma^0$ and m be an integer. If $h = O(|x|^m)$, then $\widetilde{f}_h = O(|x|^{m-1})$.*

Proof. By Proposition 3.3.5 we may suppose that $h(x) = x^m$. The lemma holds for $m \leq 1$ by Corollary 3.3.19. When $m > 1$, suppose that we have proved the lemma for $m \leq M$ where $M \in \mathbb{N}$, then for any $m \leq M + 2$, we have by Stein's equation

$$\begin{aligned}\widetilde{f}_{x^m}(x) &= x^{-1}(x^m + \sigma^2 \widetilde{f}'_{x^m}(x)) = x^{m-1} + \sigma^2 \widetilde{f}_{\Gamma(x^m)}(x)\\ &= x^{m-1} + \sigma^2(m-1)\widetilde{f}_{x^{m-2}}(x)\\ &= x^{m-1} + O(|x|^{m-3}) = O(|x|^{m-1}).\end{aligned}$$

□

The following proposition gives the order of the derivatives of $\widetilde{f}_h$. We treat the case for $\widetilde{f}_h^{(l)}$ where $l = 0, \cdots, N+1$ and the case for $\widetilde{f}_h^{(N+2)}$ differently. The former is obtained by a standard method combining Proposition 4.2.10 and 4.2.11, together with Lemma 4.2.15. For the latter, we shall use the Stein's equation.

Proposition 4.2.17 *Let m be a positive integer. Suppose $h \in \mathscr{H}_\sigma^N$ agrees with a function in $\mathscr{H}_\sigma^{N+1}$ when $|x|$ is large. Suppose in addition that $h^{(N+1)} = O(|x|^{m-1})$ and when $m = 0$, $h^{(N)}$ is bounded. Then*

1) for any integer $0 \le l \le N+1$, $\widetilde{f}_h^{(N+1-l)} = O(|x|^{m+l-2})$;

2) $\widetilde{f}_h^{(N+2)}(h) = O(|x|^{m-1})$.

Proof. Since we discuss the behavior when $|x|$ is sufficiently large, we may suppose $h \in \mathscr{H}_\sigma^{N+1}$ and then we can use $\Gamma^{N+1}(h)$ etc.
1) By (4.20) we know that

$$\widetilde{f}_h^{(N+1-l)}(x) = \sum_{k=0}^{[\frac{N+1-l}{2}]} \binom{N+1-l}{2k} (2k-1)!! x^{N+1-l-2k} \widetilde{f}_{\Gamma^{N+1-l-k}(h)}(x). \tag{4.26}$$

Since $h^{(N+1)} = O(|x|^{m-1})$ and $h^{(N)} = O(|x|^m)$ we know by Lemma 4.2.15 that for any $0 \le k \le (N+1-l)/2$, $\Gamma^{N+1-l-k}(h) = O(|x|^{m-N+2(k+l)-2})$. Therefore by Lemma 4.2.16, $\widetilde{f}_{\Gamma^{N+1-l-k}(h)} = O(|x|^{m-N+2(k+l)-3})$. So we know that $\widetilde{f}_h^{(N+1-l)}(x) = O(|x|^{m+l-2})$.

2) We let $l = 0$ in (4.26) and separate the leading order term with the others to get

$$\widetilde{f}_h^{(N+1)}(x) = x^{N+1} \widetilde{f}_{\Gamma^{N+1}(h)} + \sum_{k=1}^{[\frac{N+1}{2}]} \binom{N+1}{2k} (2k-1)!! x^{N+1-2k} \widetilde{f}_{\Gamma^{N+1-k}(h)}(x). \tag{4.27}$$

Deriving the two sides of (4.27) and making a change of indice as in the proof of the Proposition 4.2.10, we obtain

$$\widetilde{f}_h^{(N+2)}(x) = x^{N+1} \widetilde{f}'_{\Gamma^{N+1}(h)} + \sum_{k=1}^{[\frac{N+2}{2}]} \binom{N+2}{2k} (2k-1)!! x^{N+2-2k} \widetilde{f}_{\Gamma^{N+2-k}(h)}(x),$$

Since $h^{(N+1)} = O(|x|^{m-1})$ and $h^{(N)} = O(|x|^m)$ we know that for any $1 \le k \le (N+2)/2$, $\Gamma^{N+2-k}(h) = O(|x|^{m-N+2k-4})$. Therefore $\widetilde{f}_{\Gamma^{N+2-k}(h)} = O(|x|^{m-N+2k-5})$. So we know that

$$\sum_{k=1}^{[\frac{N+2}{2}]} \binom{N+2}{2k} (2k-1)!! x^{N+2-2k} \widetilde{f}_{\Gamma^{N+2-k}(h)}(x) = O(|x|^{m-3}).$$

By Stein's equation,

$$\widetilde{f}'_{\Gamma^{N+1}(h)} = \frac{1}{\sigma^2}\Big(x \widetilde{f}_{\Gamma^{N+1}(h)} - \Gamma^{N+1}(h) \Big).$$

By Lemma 4.2.16, we know that the order of $\widetilde{f}'_{\Gamma^{N+1}(h)}$ is the same as that of $\Gamma^{N+1}(h)$, which equals $m - N - 2$. So finally we get $\widetilde{f}_h^{(N+2)} = O(|x|^{m-1})$. $\square$

Remark 4.2.18 1. When $m \geq 1$, $h^{(N+1)} = O(|x|^{m-1})$ implies $h^{(N)} = O(|x|^m)$ by Lemma 4.2.15. However, this is not the case when $m = 0$. Therefore, we introduce explicitly the condition that $h^{(N)}$ is bounded if $m = 0$.

2. By using the Stein's equation in the last step, we obtain the order of $\widetilde{f}_h^{(N+2)}$ which is 2 degrees higher than what would have been obtained by using our previous procedure if there were enough regularity. This is in fact already shown by Proposition 3.3.20 where $h \in \mathscr{H}_\sigma^0$ and $h(x) = O(|x|)$, that is, $N = 0$ and $m = 1$. Then $\widetilde{f}_h^{(1-l)}(x) = O(|x|^{l-1})$ for $l = 0, 1$ and $\widetilde{f}_h'' = O(|x|^0)$.

We now specify the conditions on h in Definition 4.2.19 under which we can obtain error estimation for the N^{th} order normal expansion of $h(W)$. These conditions specify the regularity order and the growing speed order of a function h. We note that by definition of the set $\mathscr{H}_\sigma^0$, h is defined on $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$, therefore, we have to discuss the point 0 separately.

Definition 4.2.19 Let $h : \mathbb{R} \to \mathbb{R}$ be a function, N, m be positive integers. We say that h satisfies the *condition* $\mathbf{P}(N, m)$, denoted by $h \sim \mathbf{P}(N, m)$, if

1) h has up to N^{th}-order derivatives in a neighborhood of 0;

2) $h^{(N)}$ is locally of finite variation and also is $(h_{\text{cont}}^{(N)})'$ in a neighborhood of 0;

3) $h|_{\mathbb{R}_*} \in \mathscr{H}_\sigma^N$, and $h_{\text{cont}}^{(N)}\Big|_{\mathbb{R}_*} \in \mathscr{H}_\sigma^{(1)}$;

4) $h^{(N)}(x) = O(|x|^m)$ and $(h_{\text{cont}}^{(N)})' = O(|x|^{m-1})$.

In the above definition, the meaning of the integers N and m have been discussed previously. We also need some regularity conditions around the point 0 which is specified by 1) and 2).

To estimate $e(N, h)$, we should be capable to estimate $e(N - l, f_h^{(l+1)})$ for $l = 1, \cdots, N$. Therefore, we estimate by recurrence and we need to verify that $f_h^{(l+1)}$ satisfies the above conditions. The following proposition gives parameters in the conditions satisfied by f_h and its derivatives.

Proposition 4.2.20 *Let $h : \mathbb{R} \to \mathbb{R}$ be a function such that $h \sim \mathbf{P}(N, m)$.*

1) For any $m_1 \geq m$, we have $h \sim \mathbf{P}(N, m_1)$;

2) for any integer $0 \leq k \leq N$, $h^{(k)} \sim \mathbf{P}(N - k, m)$;

3) $f_h \sim \mathbf{P}(N + 1, m)$;

Proof. The first two assertions follows immediately by definition. For 3), we know by Stein's equation that $f_h' = \sigma^{-2}(xf_h - h + \Phi_\sigma(h))$. By induction we know that f_h has up to $(N+1)^{\text{th}}$ derivatives. In addition, $f_h^{(N+1)}$ is locally of finite variation, and the derivative of its continuous part is also locally of finite variation. So condition 1) and 2) in Definition 4.2.19 is satisfied.

Moreover, notice that $f_h|_{\mathbb{R}_*} = \widetilde{f}_{\bar{h}}$. Since $h \in \mathscr{H}_\sigma^N$, also is $\bar{h}$. So by Proposition 4.2.13, $\widetilde{f}_{\bar{h}} \in \mathscr{H}_\sigma^{N+1}$. Since $h_{\text{cont}}^{(N)}\big|_{\mathbb{R}_*} \in \mathscr{H}_\sigma^1$, by Stein's equation $\widetilde{f}_{\bar{h}}' = \sigma^{-2}(x\widetilde{f}_{\bar{h}} - \bar{h})$, we know that the continuous part of $\widetilde{f}_{\bar{h}}^{(N+1)}$ lies in $\mathscr{H}_\sigma^1$. Finally, by Proposition 4.2.17 and Remark 4.2.18, we know the growing speed of $f_h^{(N+1)}$ and $(f_h^{(N+1)})'_{\text{cont}}$, which follows that 4) of Definition 4.2.19 is fulfilled. □

4.2.3.4 Estimation of $e(N,h)$

We now estimate $e(N,h)$. We shall give estimations of the remaining terms δ and ε in Proposition 4.2.21 and Proposition 4.2.23 respectively. Proposition 4.2.25 gives the estimation of $e(N,h)$ in the recurrence form by summarizing the previous results.

Proposition 4.2.21 *Let g be a function satisfying the condition $\mathbf{P}(N,m)$, X and Y be two independent random variables such that $\mathbb{E}[|X|^{(m-1)^+}]$ is bounded. We suppose that c and r are two positive constants such that X verifies the concentration inequality for any real numbers $a \le b$, i.e.*

$$\mathbb{P}(a \le X \le b) \le c(b-a) + r.$$

If $0 \le k \le N$ is an integer, then $|\delta(N-k, g^{(k)}, X, Y)|$ can be bounded by a linear combination of the form

$$\sum_{j=0}^{(m-1)^+} U_j^{(k)} \mathbb{E}[|Y|^{N+1-k+j}] + rV^{(k)}\mathbb{E}[|Y|^{N-k}],$$

where $U_j^{(k)}$ is a constant which depends on g, c, k and $\mathbb{E}[|X|^{(m-1)^+-j}]$, $V^{(k)}$ is a constant which depends only on g and k.

Proof. By Taylor's formula 4.7, we have

$$\delta(N-k, g^{(k)}, X, Y) = \frac{1}{(N-k-1)!}\int_0^1 (1-t)^{N-k-1}\mathbb{E}\Big[\Big(g^{(N)}(X+tY) - g^{(N)}(X)\Big)Y^{N-k}\Big]dt.$$

Since $g \sim \mathbf{P}(N,m)$, let $u = g_{\text{cont}}^{(N)}$, $v = g^{(N)} - u$. We shall discuss the two parts separately. For the discrete part, by definition, v has finite number of jumps and is of the form $v(x) = \sum_{j=1}^M \varepsilon_j \mathbb{1}_{(-\infty,K_j]}(x)$. By (3.48), we know that $v(X+tY) - v(X) =$

$\sum_{j=1}^{M} \varepsilon_j \mathbb{1}_{\{K_j - \max(tY,0) < X \le K_j - \min(tY,0)\}}$. Moreover, by the concentration inequality hypothesis,

$$\mathbb{E}\Big[|v(X+tY)-v(X)| \;\Big|\; Y\Big] \le \sum_{j=1}^{M} |\varepsilon_j| \Big(ct|Y|+r\Big)$$

So there exist two constants A_1 and A_2 such that

$$\begin{aligned}
&\frac{1}{(N-k-1)!}\int_0^1 (1-t)^{N-k-1}\mathbb{E}\Big[\big|v(X+tY)-v(X)\big||Y|^{N-k}\Big]dt \\
&\le \frac{1}{(N-k-1)!}\Big(\sum_{j=1}^{M}|\varepsilon_j|\Big)\Big(c\int_0^1 (1-t)^{N-k-1}tdt\mathbb{E}[|Y|^{N-k+1}] + r\int_0^1 (1-t)^{N-k-1}dt\mathbb{E}[|Y|^{N-k}]\Big) \\
&\le A_1\mathbb{E}[|Y|^{N-k+1}] + A_2 r\mathbb{E}[|Y|^{N-k}]
\end{aligned}$$

where A_1 and A_2 depend on the total absolute jump size of the function g and the integer k. In addition, A_1 depends on c and A_2 depends on r. For the continuous part, u is differntiable. Then we get by comparing the remaining terms of (4.6) and (4.7)

$$\begin{aligned}
&\frac{1}{(N-k-1)!}\int_0^1 (1-t)^{N-k-1}\mathbb{E}\Big[\big(u(X+tY)-u(X)\big)Y^{N-k}\Big]dt \\
&= \frac{1}{(N-k)!}\int_0^1 (1-t)^{N-k}\mathbb{E}[u'(X+tY)Y^{N-k+1}]dt.
\end{aligned} \tag{4.28}$$

Since $u'(x) = O(|x|^{m-1})$, when $m \ge 1$, there exist two positive constants A_3 and A_4 such that

$$|u'(X+tY)| \le A_3 + A_4 \sum_{j=0}^{m-1} \binom{m-1}{j} |X|^{m-1-j}|tY|^j.$$

Then

$$\begin{aligned}
(4.28) \le \frac{1}{(N-k)!}\Big(&A_3\int_0^1 (1-t)^{N-k}dt\mathbb{E}[|Y|^{N-k+1}] \\
&+ A_4\sum_{j=0}^{m-1}\binom{m-1}{j}\int_0^1 (1-t)^{N-k}t^j dt\mathbb{E}[|X|^{m-1-j}]\mathbb{E}[|Y|^{N-k+1+j}]\Big).
\end{aligned}$$

When $m = 0$, then u is a bounded function. We have

$$(4.28) \le \frac{2\|u\|}{(N-k-1)!}\mathbb{E}[|Y|^{N-k}]\int_0^1 (1-t)^{N-k-1}dt.$$

Combining the above cases, we obtain the proposition.

$\square$

Remark 4.2.22 We shall replace X by $W^{(i)}$ and the concentration inequality is given by Corollary 3.4.5. So we know that

$$r = \frac{8\sigma_i}{\sigma_W} + \frac{2\sum_{i=1}^n \mathbb{E}\big[|X_i^s|^3\big]}{\sigma_W^3} + \frac{\Big(\sum_{i=1}^n \frac{\sigma_i}{\sqrt{2}}\mathbb{E}[|X_i^s|^3]\Big)^{\frac{1}{2}}}{\sigma_W^2}.$$

In the i.i.d. Bernoulli case, this term is of order $O(\frac{1}{\sqrt{n}})$, which ensures that $|\delta(N-k, g^{(k)}, X, Y)|$ is of the order $O\big((\frac{1}{\sqrt{n}})^{N-k+1}\big)$.

Proposition 4.2.23 *With the notation and the conditions of Proposition 4.2.21, we have*

$$|\varepsilon(N,g,X,Y)| \le \sum_{d\ge 0} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\le N}} \prod_{l=1}^d \frac{\mathbb{E}[|Y|^{j_l}]}{j_l!}\Big(\sum_{i=0}^{(m-1)^+} U_i^{(|J|)}\mathbb{E}[|Y|^{N+1-|J|+i}] + rV^{(|J|)}\mathbb{E}[|Y|^{N-|J|}]\Big), \tag{4.29}$$

where $U_i^{(k)}$ and $V^{(k)}$ are the constants in Proposition 4.2.21.

Proof. It is a direct consequence of Proposition 4.2.21 and Proposition 4.2.2. □

Remark 4.2.24 Proposition 4.2.23 shows that $|\varepsilon(N,g,X,Y)|$ is of order $O\big((\frac{1}{\sqrt{n}})^{N+1}\big)$.

Using Proposition 4.2.21 and Proposition 4.2.23, we can obtain an upper estimation bound of $e(N,h)$ given by (4.13). The following result shows that it is of the correct order in the binomial case.

Proposition 4.2.25 *Let $h \sim \mathbf{P}(N,m)$. Suppose that $X_1, \cdots, X_n$ are i.i.d Bernoulli random variables and that $W = X_1 + \cdots + X_n$ is of finite variance, i.e. $\sigma_W < \infty$. Then*

$$e(N,h) \sim O\left(\Big(\frac{1}{\sqrt{n}}\Big)^{N+1}\right).$$

Proof. We shall prove by deduction. When $N=0$, $h \sim \mathbf{P}(0,m)$. We have $|e(0,h)| = |\mathbb{E}[h(W)] - \Phi_{\sigma_W}(h)| = \sigma_W^2 \mathbb{E}\big[|f_h'(W^*) - f_h'(W)|\big]$. By Proposition 4.2.20, $f_h \sim \mathbf{P}(1,m)$. Let $u = (f_h')_{\text{cont}}$ and $v = f_h' - u$. Then $|e(0,h)| \le \mathbb{E}\big[|u(W^*) - u(W)|\big] + \mathbb{E}\big[|v(W^*) - v(W)|\big]$. From Proposition 3.4.6, we know that $\mathbb{E}\big[|v(W^*) - v(W)|\big] \sim O\big(\frac{1}{\sqrt{n}}\big)$. On the other hand, $u' = O(|x|^{m-1})$,

$$\mathbb{E}\big[|u(W^*) - u(W)|\big] \le \int_0^1 \mathbb{E}\big[|u'(W^{(i)} + tX_1^*)X_1^*|\big] + \mathbb{E}\big[|u'(W^{(i)} + tX_1)X_1|\big]dt \sim O\Big(\frac{1}{\sqrt{n}}\Big).$$

Suppose that we have proved for $0, 1, \cdots, N-1$. Since $h \sim \mathbf{P}(N,m)$, we have $f_h \sim \mathbf{P}(N+1,m)$, which implies that $f_h' \sim \mathbf{P}(N,m)$ and $f_h^{(k+1)} \sim \mathbf{P}(N-k,m)$. We apply

Proposition 4.2.21 and Proposition 4.2.23 to obtain that $\sum_{i=1}^n \sigma_i^2 \sum_{k=0}^N \frac{\mathbb{E}[(X_i^*)^k]}{k!} \varepsilon\big(N - k, f_h^{(k+1)}, W^{(i)}, X_i\big)$ and $\sum_{i=1}^n \sigma_i^2 \delta\big(N, f_h', W^{(i)}, X_i^*\big)$ in (4.13) is of order $O\big((\frac{1}{\sqrt{n}})^{N+1}\big)$. Moreover, we have by recurrence that $e\big(N - |\mathbf{J}|, f_h^{(|\mathbf{J}|+1)}\big) \sim O\big((\frac{1}{\sqrt{n}})^{|J|+1}\big)$, which implies that $e(N, h)$ is of order $O\big((\frac{1}{\sqrt{n}})^{N+1}\big)$. $\square$

4.3 Poisson approximation

This section deals with the asymptotic expansion of $\mathbb{E}[h(W)]$ by the Poisson approximation. We shall show that our method can be adapted without any difficulty in the Poisson case. The results obtained are very similar with those of the previous section.

4.3.1 Preliminaries

4.3.1.1 Framework

Chen [15] observes that a $\mathbb{N}^+$-valued (non-negative) random variable Z follows the Poisson distribution, i.e. $Z \sim P(\lambda)$ if and only if

$$\mathbb{E}[Zf(Z)] = \lambda \mathbb{E}[f(Z+1)]$$

for any bounded function f. This similarity with the normal case motivates us to define the zero bias transformation in the Poisson case. In the following of this section, Z represents a Poisson random variable.

Definition 4.3.1 Let X be a random variable taking non-negative integer values and $\mathbb{E}[X] = \lambda < \infty$. Then X^* is said to have the *X-Poisson zero biased distribution* if for any function f such that $\mathbb{E}[Xf(X)]$ exists, we have

$$\mathbb{E}[Xf(X)] = \lambda \mathbb{E}[f(X^* + 1)]. \tag{4.30}$$

Example 4.3.2 Let X be a Bernoulli random variable with $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = q = 1 - p$. Then

$$\frac{1}{p}\mathbb{E}[Xf(X)] = \frac{1}{p}\big(pf(1)\big) = f(1),$$

which means that X^* exists and has the Dirac distribution δ_0.

Remark 4.3.3 We here consider the standard Bernoulli random variables instead of the asymmetric ones.

Proposition 4.3.4 *Let X be a random variable taking non-negative integer values with finite expectation $\lambda > 0$. Then there exists X^* which has the X-Poisson zero bias distribution. Moreover, the distribution of X^* is unique and is given by*

$$\mathbb{P}(X^* = a) = \frac{a+1}{\lambda}\mathbb{P}(X = a+1), \tag{4.31}$$

Proof. We first prove the uniqueness. Let $f(x) = \mathbb{1}_{\{x=a+1\}}$ in (4.30) where $a \in \mathbb{N}$, then we have

$$\lambda\mathbb{P}(X^* = a) = (a+1)\mathbb{P}(X = a+1).$$

So the distribution of X^* is uniquely determined.
Let X^* be a random variable satisfying (4.31). Then for any $\mathbb{N}^+$-valued function f such that $\mathbb{E}[Xf(X)]$ exists, we have

$$\begin{aligned}\frac{1}{\lambda}\mathbb{E}[Xf(X)] &= \frac{1}{\lambda}\sum_{a=0}^{\infty}\mathbb{P}(X = a)af(a)\\ &= \frac{1}{\lambda}\sum_{a=0}^{\infty}\mathbb{P}(X = a+1)(a+1)f(a+1) = \mathbb{E}[f(X^*+1)].\end{aligned}$$

So X^* has the X-Poisson zero biased distribution. □

Proposition 4.3.5 *1. For any integer $k \geq 1$,*

$$\mathbb{E}[(X^*)^k] = \frac{1}{\lambda}\mathbb{E}[X(X-1)^k]. \tag{4.32}$$

In particular, $\mathbb{E}[X^*] = \mathbb{E}[X^2] - 1$.

2. Let X^* be a random variable which has X-Poisson zero bias distribution and is independent with X. Then for any function f which takes non negative integer values such that $\mathbb{E}[|f(X^* - X)|] < +\infty$,

$$\mathbb{E}[f(X^* - X)] = \frac{1}{\lambda}\mathbb{E}[Xf(X^s - 1)] \tag{4.33}$$

where $X^s = X - \widetilde{X}$ and $\widetilde{X}$ is an independent copy of X. In particular, for any integer $k \geq 1$ such that $\mathbb{E}[|X^* - X|^k] < +\infty$,

$$\mathbb{E}\big[|X^* - X|^k\big] = \frac{1}{\lambda}\mathbb{E}\big[X|X^s - 1|^k\big]. \tag{4.34}$$

Proof. 1) Let $f(x) = (x-1)^k$, Then (4.32) follows immediately by definition.
2) For any $a \in \mathbb{N}$, we have by (4.30)

$$\mathbb{E}[f(X^* - a)] = \frac{1}{\lambda}\mathbb{E}[Xf(X - a - 1)].$$

Since X^* and X are independent, we know that

$$\mathbb{E}[f(X^* - X)] = \frac{1}{\lambda}\mathbb{E}[Xf(X^s - 1)],$$

where $\widetilde{X}$ is an independent copy of X. (4.34) is then a direct consequence. $\square$

The Poisson zero bias transformation for the sum of independent random variables is given in the same way as in the normal case.

Proposition 4.3.6 *Let $X_1 . \cdots , X_n$ be independent random variables with positive expectations $\lambda_1, \cdots , \lambda_n$. Denote by $W = X_1 + \cdots + X_n$ and $\lambda_W = \mathbb{E}[W]$. Let I be a random index independent of X_i satisfying*

$$P(I = i) = \lambda_i/\lambda_W.$$

Let X_i^* be a random variable having the X_i-Poisson zero bias distribution and independent of all X_j and I. Then $W^{(I)} + X_I^*$ have the W-Poisson zero bias distribution where $W^{(i)} = W - X_i$.

Proof. The proof is the same with that of Proposition 3.2.10 by replacing the normal zero bias transformation with the Poisson one. $\square$

Corollary 4.3.7 *With the notation of Proposition 4.3.6, we have*

$$\mathbb{E}[|W^* - W|^k] = \frac{1}{\lambda_W}\sum_{i=1}^{n}\mathbb{E}[X_i|X_i^s - 1|^k]. \tag{4.35}$$

Proof. (4.35) is direct by (4.34) and Proposition 4.3.6. $\square$

In the following, we denote by $\mathcal{P}_\lambda(h) = \mathbb{E}[h(Z)]$ where $Z \sim P(\lambda)$. For any $\mathbb{N}^+$-valued function h such that $\mathcal{P}_\lambda(h)$ is well defined, we introduce the *Stein's Poisson equation* given by Chen [15] as below:

$$xp(x) - \lambda p(x+1) = h(x) - \mathcal{P}_\lambda(h). \tag{4.36}$$

where p is an auxiliary function. The solution of (4.36) is given by

$$p(a) = \frac{(a-1)!}{\lambda^a}\sum_{i=a}^{\infty}\frac{\lambda^i}{i!}\big(h(i) - \mathcal{P}_\lambda(h)\big). \tag{4.37}$$

It is unique except at $a = 0$. However, the value $p(0)$ does not enter into our calculation afterwards. There exists a recurrence form of the solution given by

$$p(1) = \frac{\mathcal{P}_\lambda(h) - h(0)}{\lambda}, \cdots , p(a+1) = \frac{\mathcal{P}_\lambda(h) - h(a) + ap(a)}{\lambda}.$$

In the following, we denote by $p_{h,\lambda}$ or simply by p_h when there is no ambiguity the solution (4.37). Combining (4.30) and (4.36), we obtain, for any random variable W with expectation $\lambda_W \leq \infty$,

$$\mathbb{E}[h(W)] - \mathcal{P}_{\lambda_W}(h) = \lambda_W \mathbb{E}[p_h(W^* + 1) - p_h(W + 1)]. \tag{4.38}$$

4.3.1.2 First order estimation

We obtain immediately a first order estimation under this framework. In fact, Le Cam (1960) showed that for independent Bernoulli random variables $X_1, \cdots, X_n$ with $\mathbb{P}(X_i = 1) = p_i$ and $\mathbb{P}(X_i = 0) = 1 - p_i$, we have

$$\left|\mathbb{E}[h(W)] - \mathcal{P}_{\lambda_W}\right| \leq 2\|h\| \sum_{i=1}^{n} p_i^2$$

where $W = \sum_{i=1}^{n} X_i$ and $\lambda_W = \sum_{i=1}^{n} p_i$. Chen [15] used the Stein's method to obtain a similar result where $2\|h\|$ is replaced with $6\|h\| \min\big((\lambda_W)^{-\frac{1}{2}}, 1\big)$ since he proved that $\|\Delta p_h\| \leq 6\|h\| \min\big((\lambda_W)^{-\frac{1}{2}}, 1\big)$.

Combining (4.38) and (4.35), we obtain immediately that for any $\mathbb{N}^+$-valued random variables $X_1, \cdots, X_n$ and any $\mathbb{N}^+$-valued function h,

$$\begin{aligned}\left|\mathbb{E}[h(W)] - \mathcal{P}_{\lambda_W}(h)\right| &\leq \lambda_W \|\Delta p_h\| \mathbb{E}\big[|W^* - W|\big] \\ &\leq 6\|h\| \min\Big(\frac{1}{\sqrt{\lambda_W}}, 1\Big) \sum_{i=1}^{n} \mathbb{E}\big[X_i |X_i - 1 - \widetilde{X}_i|\big]\end{aligned} \tag{4.39}$$

where $\widetilde{X}_i$ is an independent duplicate of X_i. In particular, if X_i is a Bernoulli random variable of parameter $\lambda_i = p_i$, then $\mathbb{E}\big[|W^* - W|\big] = \frac{1}{\lambda_W} \sum_{i=1}^{n} p_i^2$, which corresponds to the result of Le Cam and Chen.

4.3.1.3 Some properties in the discrete case

We now present some useful results in the Poisson calculation. On one hand, they are comparable to those in the Gaussian case. On the other hand, the techniques used are very different. We first recall the expansion formula of the difference method, which is analogous with the Taylor expansion in the continuous case. In the following, we denote by $\Delta p(x) = p(x + 1) - p(x)$.

Proposition 4.3.8 *For any integer $k \geq 1$ and $m \geq 1$, we have*

$$p(x + k) = \sum_{j=0}^{m} \binom{k}{j} \Delta^j p(x) + \sum_{0 \leq j_1 < \cdots < j_{m+1} < k} \Delta^{m+1} p(x + j_1). \tag{4.40}$$

In particular,

$$p(x+k) = \sum_{j=0}^{k} \binom{k}{j} \Delta^j p(x) = (1+\Delta)^k p(x). \tag{4.41}$$

Proof. We shall prove by induction. When $m = 0$, (4.40) holds since

$$p(x+k) - p(x) = \sum_{j=0}^{k-1} \Delta p(x+j).$$

Suppose we have proved (4.40) for $0, 1, \cdots, m-1$, then

$$\begin{aligned}
p(x+k) &= \sum_{j=0}^{m-1} \binom{k}{j} \Delta^j f(x) + \sum_{0 \le j_2 < \cdots < j_{m+1} < k} \Delta^m p(x+j_2) \\
&= \sum_{j=0}^{m-1} \binom{k}{j} \Delta^j p(x) + \sum_{0 \le j_2 < \cdots < j_{m+1} < k} \Big(\Delta^m p(x) + \sum_{j_1=0}^{j_2-1} \Delta^{m+1} p(x+j_1) \Big) \\
&= \sum_{j=0}^{m} \binom{k}{j} \Delta^j p(x) + \sum_{0 \le j_1 < \cdots < j_{m+1} < k} \Delta^{m+1} p(x+j_1).
\end{aligned}$$

In particular, if $m > k$, $\{0 \le j_1 < \cdots j_{m+1} < k\}$ is the empty set and the summation term equal zero, which follows (4.41). $\square$

In the discrete case, the difference Δ replaces the derivative, and in the place of calculating the normal expectation for functions of the form $x^m p_h^{(l)}$, we are interested in writing the Poisson expectation for functions of the form $\binom{x}{m} \Delta^l p_h(x)$ as that for a polynomial function of $h(x)$. We first recall two simple results.

Lemma 4.3.9 *1. For all positive integers a and b with $a > b$, we have*

$$\binom{a}{b} = \binom{a-1}{b} + \binom{a-1}{b-1}. \tag{4.42}$$

2. For any functions f and g,

$$\Delta(f(x)g(x)) = f(x+1)\Delta g(x) + (\Delta f(x))g(x). \tag{4.43}$$

Proof. Direct calculations give immediately (4.42) and (4.43). $\square$

Proposition 4.3.10 *For integers $m \ge 1$ and $l \ge 0$, suppose that*

$$\mathcal{P}_\lambda\Big(\binom{x}{m} \Delta^l p_h(x)\Big) = \mathcal{P}_\lambda\big(P_{m,l}(\Delta) h(x)\big),$$

then we have the recurrence form

$$P_{m,0}(z) = \frac{\lambda^m}{m \cdot m!}\big((1+z)^m - 1\big), \qquad P_{m,l} = \frac{m+1}{\lambda} P_{m+1,l-1} - P_{m,l-1}, \tag{4.44}$$

and the explicit form

$$P_{m,l}(z) = \frac{\lambda^m}{m!} \sum_{i=0}^{l} \binom{l}{i} \frac{(-1)^{l-i}}{m+i} ((1+z)^{m+i} - 1). \tag{4.45}$$

In particular, $P_{1,l}(z) = \frac{\lambda}{l+1} z^{l+1}$.

Proof. When $l = 0$, we have by definition and (4.37)

$$\begin{aligned}
\mathcal{P}_\lambda\left(\binom{x}{m} p_h(x)\right) &= \sum_{k \geq m} e^{-\lambda} \frac{\lambda^k}{k!} \binom{k}{m} \frac{(k-1)!}{\lambda^k} \sum_{i=k}^{\infty} \frac{\lambda^i}{i!} \big(h(i) - \mathcal{P}_\lambda(h)\big) \\
&= \frac{e^{-\lambda}}{m!} \sum_{k \geq m} \frac{(k-1)!}{(k-m)!} \sum_{i=k}^{\infty} \frac{\lambda^i}{i!} (h(i) - \mathcal{P}_\lambda(h)) \\
&= \frac{e^{-\lambda}}{m!} \sum_{i=m}^{\infty} \sum_{k=m}^{i} \frac{(k-1)!}{(k-m)!} \frac{\lambda^i}{i!} \big(h(i) - \mathcal{P}_\lambda(h)\big).
\end{aligned}$$

By (4.42),

$$\begin{aligned}
\sum_{k=m}^{i} \frac{(k-1)!}{(k-m)!} &= \sum_{k=0}^{i-m} \frac{(k+m-1)!}{k!} = (m-1)! \sum_{k=0}^{i-m} \binom{k+m-1}{m-1} \\
&= (m-1)! \sum_{k=0}^{i-m} \binom{k+m}{m} - \binom{k+m-1}{m} = (m-1)! \binom{i}{m}.
\end{aligned}$$

Then

$$\begin{aligned}
&\mathcal{P}_\lambda\Big(\binom{x}{m} p_h(x)\Big) \\
&= \frac{e^{-\lambda}}{m!} \sum_{i=m}^{\infty} (m-1)! \binom{i}{m} \frac{\lambda^i}{i!} \big(h(i) - \mathcal{P}_\lambda(h)\big) = \frac{e^{-\lambda}}{m \cdot m!} \sum_{i=m}^{\infty} \frac{\lambda^i}{(i-m)!} (h(i) - \mathcal{P}_\lambda(h)) \\
&= \frac{e^{-\lambda}}{m \cdot m!} \sum_{i=0}^{\infty} \frac{\lambda^{i+m}}{i!} (h(i+m) - \mathcal{P}_\lambda(h)) = \frac{\lambda^m}{m \cdot m!} \big(\mathcal{P}_\lambda(h(x+m)) - \mathcal{P}_\lambda(h)\big) \\
&= \mathcal{P}_\lambda\Big(\frac{\lambda^m}{m \cdot m!} \big((1+\Delta)^n h(x) - h(x)\big)\Big).
\end{aligned}$$

Hence

$$P_{m,0}(z) = \frac{\lambda^m}{m \cdot m!} \big((1+z)^m - 1\big).$$

We verify that both (4.44) and (4.45) hold for $l = 0$. We now consider the case when $l > 0$. Denote by

$$\Delta\binom{x}{m} = \binom{x+1}{m} - \binom{x}{m} = \binom{x}{m-1},$$

then we get by (4.43)

$$\binom{x}{m}\Delta^l p(x) = \Delta\left(\binom{x-1}{m}\Delta^{l-1}p(x)\right) - \Delta\binom{x-1}{m}\Delta^{l-1}p(x),$$

On the other hand, by the invariant property of the Poisson distribution under the zero bias transformation, we have

$$\mathcal{P}_\lambda(g(x+1)) = \frac{1}{\lambda}\mathcal{P}_\lambda(xg(x)), \tag{4.46}$$

which implies

$$\mathcal{P}_\lambda(\Delta g) = \mathcal{P}_\lambda\left(\left(\frac{x}{\lambda}-1\right)g(x)\right).$$

Then

$$\mathcal{P}_\lambda\left(\binom{x}{m}\Delta^l p(x)\right) = \mathcal{P}_\lambda\left(\left(\frac{x}{\lambda}-1\right)\binom{x-1}{m}\Delta^{l-1}p\right) - \mathcal{P}_\lambda\left(\binom{x-1}{m-1}\Delta^{l-1}p\right)$$

Then it suffice to apply again (4.42) to get

$$\begin{aligned}\mathcal{P}_\lambda\left(\binom{x}{m}\Delta^l p(x)\right) &= \frac{1}{\lambda}\mathcal{P}_\lambda\left(x\frac{(x-1)!}{m!(x-1-m)!}\Delta^{l-1}p\right) - \mathcal{P}_\lambda\left(\binom{x}{m}\Delta^{l-1}p\right)\\ &= \frac{m+1}{\lambda}\mathcal{P}_\lambda\left(\binom{x}{m+1}\Delta^{l-1}p\right) - \mathcal{P}_\lambda\left(\binom{x}{m}\Delta^{l-1}p\right)\end{aligned}$$

which means $P_{m,l} = \frac{m+1}{\lambda}P_{m+1,l-1} - P_{m,l-1}$. which proves (4.44). To prove (4.45), we deduce by induction. Suppose (4.45) is verified for $0, 1, \cdots, l-1$. Replacing $P_{m+1,l-1}$ and $P_{m,l-1}$ in (4.44) by (4.45), we get

$$\begin{aligned}&P_{m,l}(z)\\ &= \frac{\lambda^m}{m!}\sum_{i=0}^{l-1}(-1)^{l-1-i}\binom{l-1}{i}\left(\frac{(1+z)^{m+1+i}-1}{m+1+i} - \frac{(1+z)^{m+i}-1}{m+i}\right) \qquad (4.47)\\ &= \frac{\lambda^m}{m!}\left(\sum_{i=1}^{l-1}\binom{l}{i}\frac{(-1)^{l-i}}{m+i}\big((1+z)^{m+i}-1\big) + \sum_{i=0}^{l-1}\binom{l-1}{i}\frac{(-1)^{l-1-i}}{m+i+1}\big((1+z)^{m+1+i}-1\big)\right.\\ &\left.+\sum_{i=1}^{l}\binom{l-1}{i-1}\frac{(-1)^{l-1-i}}{m+i}\big((1+z)^{m+i}-1\big)\right)\end{aligned} \tag{4.48}$$

The last equality is obtained by applying (4.42) to the second term of (4.47). At last, notice that most terms cancel each other in the second and third terms of (4.48) and we get (4.45). In particular, when $m = 1$,

$$P_{1,l}(z) = \lambda \sum_{i=0}^{l} \binom{l}{i} \frac{(-1)^{l-i}}{i+1} \big((1+z)^{i+1} - 1\big)$$
$$= \frac{\lambda}{l+1} \sum_{i=0}^{l} \binom{l+1}{i+1} (-1)^{l-i} \big((1+z)^{i+1} - 1\big) = \frac{\lambda}{l+1} z^{l+1},$$

which ends the proof. □

4.3.2 Asymptotic expansion for Bernoulli random variables

In this subsection, we consider the case where $X_1, \cdots, X_n$ are independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = p_i$ and $\mathbb{P}(X_i = 0) = 1 - p_i$. Then the expectation $\lambda_i = p_i$ and the Poisson zero bias transformation X_i^* follows Dirac distribution.

Lemma 4.3.11 *For any integer number $m \geq 0$ and any function h such that (4.49) is well defined, we have*

$$\mathbb{E}\big[h(W^{(i)})\big] = \sum_{j=0}^{m} \lambda_i^j (-1)^j \mathbb{E}\big[\Delta^j h(W)\big] + (-1)^{m+1} \lambda_i^{m+1} \mathbb{E}\big[\Delta^{m+1} h(W^{(i)})\big]. \qquad (4.49)$$

Proof. Since X_i is independent of $W^{(i)}$,

$$\mathbb{E}[h(W^{(i)} + X_i)] = p_i \mathbb{E}[h(W^{(i)} + 1)] + (1 - p_i)\mathbb{E}[h(W^{(i)})],$$

which follows

$$\mathbb{E}[h(W^{(i)})] = \mathbb{E}[h(W)] - p_i \mathbb{E}[\Delta h(W^{(i)})]. \qquad (4.50)$$

So (4.49) holds for $m = 0$. Suppose we have proved for $0, 1, \cdots, m-1$, then we apply (4.50) to the term $\mathbb{E}[\Delta^m h(W^{(i)})]$ to obtain that (4.49) holds for m. □

Proposition 4.3.12 *Let $\mathbb{E}[h(W)] = C(N,h) + e(N,h)$, then*

$$C(N,h) = C(0,h) + \sum_{i=1}^{n} \sum_{j=1}^{N} (-1)^j \lambda_i^{j+1} C\big(N - j, \Delta^j p_h(x+1)\big) \qquad (4.51)$$

where $C(0,h) = \mathcal{P}_{\lambda_W}(h)$ and

$$e(N,h) = \sum_{i=1}^{n} \sum_{j=1}^{N} (-1)^j \lambda_i^{j+1} e\big(N-j, \Delta^j p_h(x+1)\big) + (-1)^{N+1} \sum_{i=1}^{n} \lambda_i^{N+2} \mathbb{E}\big[\Delta^{N+1} p_h(W^{(i)}+1)\big]. \qquad (4.52)$$

Proof. Since X_i^* follows the Dirac distribution δ_0 and is independent of $W^{(i)}$,

$$\lambda_W \mathbb{E}\big[p_h(W^*+1) - p_h(W+1)\big] = \sum_{i=1}^n \lambda_i \mathbb{E}\big[p_h(W^{(i)}+1) - p_h(W+1)\big].$$

Applying Lemma 4.3.11 to $\mathbb{E}[p_h(W^{(i)}+1)]$, the first term when $j=0$ cancels with $\mathbb{E}[p_h(W+1)]$, so we get

$$\begin{aligned}
&\mathbb{E}[h(W)] - \mathcal{P}_{\lambda_W}(h) \\
&= \sum_{i=1}^n \sum_{j=1}^N (-1)^j \lambda_i^{j+1} \mathbb{E}\big[\Delta^j p_h(W+1)\big] + (-1)^{N+1} \sum_{i=1}^n \lambda_i^{N+2} \mathbb{E}\big[\Delta^{N+1} p_h(W^{(i)}+1)\big].
\end{aligned}$$

We then replace $\mathbb{E}\big[\Delta^j p_h(W+1)\big]$ by its $(m-j)^{\text{th}}$ order expansion to get (4.51) and (4.52). $\square$

Corollary 4.3.13 *The first two orders expansions are given as follows.*

$$C(1,h) = \mathcal{P}_{\lambda_W}(h) - \frac{1}{2}\big(\sum_{i=1}^n \lambda_i^2\big)\mathcal{P}_{\lambda_W}(\Delta^2 h)$$

and

$$\begin{aligned}
C(2,h) = {} & \mathcal{P}_{\lambda_W}(h) - \frac{1}{2}\big(\sum_{i=1}^n \lambda_i^2\big)\mathcal{P}_{\lambda_W}(\Delta^2 h) + \frac{1}{3}\big(\sum_{i=1}^n \lambda_i^3\big)\mathcal{P}_{\lambda_W}(\Delta^3 h) \\
& + \frac{1}{8}\big(\sum_{i=1}^n \lambda_i^2\big)^2 \mathcal{P}_{\lambda_W}(\Delta^4 h).
\end{aligned}$$

Proof. 1) By (4.51), $C(1,h) = \mathcal{P}_{\lambda_W}(h) - \big(\sum_{i=1}^n \lambda_i^2\big)\mathcal{P}_{\lambda_W}\big(\Delta p_h(x+1)\big)$. Combining (4.46) and Proposition 4.3.10 with $m=l=1$ follows

$$\mathcal{P}_{\lambda_W}\big(\Delta p_h(x+1)\big) = \frac{1}{\lambda_W}\mathcal{P}_{\lambda_W}\big(x\Delta p_h(x)\big) = \frac{1}{2}\mathcal{P}_{\lambda_W}(\Delta^2 h).$$

2) The calculation is similar. (4.51) yields

$$C(2,h) = \mathcal{P}_{\lambda_W}(h) - \big(\sum_{i=1}^n \lambda_i^2\big)C(1, \Delta p_h(x+1)) + \big(\sum_{i=1}^n \lambda_i^3\big)C(0, \Delta^2 p_h(x+1)).$$

Then it suffices to calculate

$$\begin{aligned}
C(1, \Delta p_h(x+1)) &= \mathcal{P}_{\lambda_W}(\Delta p_h(x+1)) - \frac{1}{2}\big(\sum_{i=1}^n \lambda_i^2\big)\mathcal{P}_{\lambda_W}(\Delta p_{\Delta p_h(x+1)}(x+1)) \\
&= \frac{1}{2}\mathcal{P}_{\lambda_W}(\Delta^2 h(x)) - \frac{1}{8}\big(\sum_{i=1}^n \lambda_i^2\big)\mathcal{P}_{\lambda_W}(\Delta^4 h)
\end{aligned}$$

and $C(0, \Delta^2 p_h(x+1)) = \frac{1}{3}\mathcal{P}_{\lambda_W}(\Delta^3 h)$. $\square$

The asymptotic expansion in the $0-1$ case has been discussed by many authors. The first two orders expansions given in Corollary 4.3.13 corresponds to those in Barbour, Chen and Choi [5]. The second and higher orders expansion have been obtained by Barbour [4] and Borisov and Ruzankin [11].

4.3.3 The general case

This subsection deals with the general case where $X_1, \cdots, X_n$ are $\mathbb{N}^+$-valued random variables. Similar as in the normal case, we introduce the following notations. Let X and Y be two independent $\mathbb{N}^+$-valued random variables and p any function on non-negative integers. Then we denote by $\delta(N, p, X, Y)$ the remaining term of the N^{th}-order difference expansion of $\mathbb{E}[p(X+Y)]$

$$\mathbb{E}[p(X+Y)] = \sum_{k=0}^{N} \mathbb{E}\Big[\binom{Y}{k}\Big]\mathbb{E}[\Delta^k p(X)] + \delta(N, p, X, Y), \tag{4.53}$$

which implies, by (4.40), that

$$\delta(N, p, X, Y) = \mathbb{E}\big[\sum_{0 \le j_1 < \cdots < j_{N+1} < Y} \Delta^{N+1} p(X + j_1)\big]. \tag{4.54}$$

The following expansion (4.55) gives the reversed Taylor's formula in the Poisson case, which enables us to write $\mathbb{E}[p(W^{(i)})]$ as an expansion on W. The relationship between ε et δ is also given below. Compared to (4.8) in the normal case, the form differs only slightly:

1) the differences replace the derivatives of the same order;

2) the terms $\mathbb{E}\big[\binom{Y}{j_l}\big]$ replace $\frac{\mathbb{E}[Y^{j_l}]}{j_l!}$.

Proposition 4.3.14 *Let N be a positive integer. Let $\varepsilon(N, p, X, Y)$ be the remaining term of the following expansion*

$$\begin{aligned}\mathbb{E}[p(X)] = \mathbb{E}[p(X+Y)] + \sum_{d \ge 1} (-1)^d \sum_{\substack{\mathbf{J}=(j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \le N}} \mathbb{E}[\Delta^{|\mathbf{J}|} p(X+Y)] \prod_{l=1}^{d} \mathbb{E}\Big[\binom{Y}{j_l}\Big] \\ + \varepsilon(N, p, X, Y)\end{aligned} \tag{4.55}$$

where for any $\mathbf{J} = (j_l) \in \mathbb{N}_*^d$, $|\mathbf{J}| = j_1 + \cdots + j_d$. Then

$$\varepsilon(N, p, X, Y) = -\sum_{d \ge 0} (-1)^d \sum_{\substack{\mathbf{J}=(j_l) \in \mathbb{N}_*^d \\ |\mathbf{J}| \le N}} \delta(N - |\mathbf{J}|, \Delta^{|\mathbf{J}|} p, X, Y) \prod_{l=1}^{d} \mathbb{E}\Big[\binom{Y}{j_l}\Big]. \tag{4.56}$$

Proof. By definition of (4.53) and (4.55),

$$\begin{aligned}\varepsilon(N,p,X,Y) = -\delta(N,p,X,Y) - \sum_{k=1}^{N}\mathbb{E}[\Delta^k p(X)]\mathbb{E}\Big[\binom{Y}{k}\Big] \\ - \sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}|}p(X+Y)]\prod_{l=1}^{d}\mathbb{E}\Big[\binom{Y}{j_l}\Big].\end{aligned} \tag{4.57}$$

On the other hand, taking Taylor expansion of the last term in the above equality gives

$$\begin{aligned}&\sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}|}p(X+Y)]\prod_{l=1}^{d}\mathbb{E}\left[\binom{Y}{j_l}\right] \\ &= \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} \left(\sum_{k=0}^{N-|\mathbf{J}|}\mathbb{E}[\Delta^{|\mathbf{J}|+k}p(X)]\mathbb{E}\left[\binom{Y}{k}\right] + \delta(N-|\mathbf{J}|,\Delta^{|\mathbf{J}|}p,X,Y)\right)\prod_{l=1}^{d}\mathbb{E}\left[\binom{Y}{j_l}\right] \end{aligned} \tag{4.58}$$

$$\begin{aligned}&= \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}|}p(X)]\prod_{l=1}^{d}\mathbb{E}\left[\binom{Y}{j_l}\right] + \sum_{\substack{\mathbf{J}'=(j_l')\in\mathbb{N}_*^{d+1}\\ |\mathbf{J}'|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}'|}p(X)]\prod_{l=1}^{d+1}\mathbb{E}\left[\binom{Y}{j_l}\right] \\ &+ \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} \delta(N-|\mathbf{J}|,\Delta^{|\mathbf{J}|}p,X,Y)\prod_{l=1}^{d}\mathbb{E}\left[\binom{Y}{j_l}\right].\end{aligned} \tag{4.59}$$

The equality (4.59) is obtained by writing respectively the term when $k=0$ in (4.58) and the summation term when $k\geq 1$. Multiplying $(-1)^d$ by (4.59) and taking the sum on d, most terms cancel and we get

$$\begin{aligned}&\sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}|}p(X+Y)]\prod_{l=1}^{d}\mathbb{E}\left[\binom{Y}{j_l}\right] \\ &= \sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} \delta(N-|\mathbf{J}|,\Delta^{|\mathbf{J}|}p,X,Y)\prod_{l=1}^{d}\mathbb{E}\left[\binom{Y}{j_l}\right] - \sum_{j\leq N}\mathbb{E}[\Delta^j p(X)]\mathbb{E}\left[\binom{Y}{j}\right],\end{aligned}$$

which follows (4.56) by noting $\sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^0\\ |\mathbf{J}|\leq N}} \delta(N-|\mathbf{J}|,\Delta^{|\mathbf{J}|}p,X,Y) = \delta(N,p,X,Y)$. □

Proposition 4.3.15 *With the notation of (4.53) and (4.55), we have*

1.

$$|\delta(N,p,X,Y)| \leq \|\Delta^{N+1}p\|\mathbb{E}\left[\binom{Y}{N+1}\right].$$

2.

$$|\varepsilon(N,p,X,Y)| \leq \ |\Delta^{N+1}p\| \sum_{d\geq 1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|=N+1}} \prod_{l=1}^{d} \mathbb{E}\left[\binom{Y}{j_l}\right].$$

Proof. 1) is obvious by definition.
2) By (4.56) and 1),

$$\begin{aligned}|\varepsilon(N,p,X,Y)| &\leq \sum_{d\geq 0} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} \|\Delta^{N+1}p\|\mathbb{E}\left[\binom{Y}{N-|\mathbf{J}|+1}\right]\left(\prod_{l=1}^{d}\mathbb{E}\left[\binom{Y}{j_l}\right]\right) \\ &\leq \|\Delta^{N+1}p\| \sum_{d\geq 1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|=N+1}} \prod_{l=1}^{d}\mathbb{E}\left[\binom{Y}{j_l}\right].\end{aligned}$$

□

The following theorem is also similar with Theorem 4.2.5, both in form and in the proof method.

Theorem 4.3.16 *For any integer $N \geq 0$, let $\mathbb{E}[h(W)] = C(N,h) + e(N,h)$ with $C(0,h) = \mathcal{P}_{\lambda_W}(h)$ and $e(0,h) = \mathbb{E}[h(W)] - \mathcal{P}_{\lambda_W}(h)$. Then*

$$\begin{aligned}C(N,h) = \mathcal{P}_{\lambda_W}(h) + \sum_{i=1}^{n}\lambda_i \sum_{d\geq 1}(-1)^{d-1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} C(N-|\mathbf{J}|, \Delta^{|\mathbf{J}|}p_h(x+1)) \\ \prod_{l=1}^{d-1}\mathbb{E}\left[\binom{X_i}{j_l}\right]\mathbb{E}\left[\binom{X_i^*}{j_d} - \binom{X_i}{j_d}\right],\end{aligned} \tag{4.60}$$

and for any $N \geq 1$,

$$\begin{aligned}&e(N,h) \\ &= \sum_{i=1}^{n}\lambda_i \sum_{d\geq 1}(-1)^{d-1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} e(N-|\mathbf{J}|, \Delta^{|\mathbf{J}|}p_h(x+1)) \prod_{l=1}^{d-1}\mathbb{E}\left[\binom{X_i}{j_l}\right]\mathbb{E}\left[\binom{X_i^*}{j_d} - \binom{X_i}{j_d}\right] \\ &+ \sum_{i=1}^{n}\lambda_i \sum_{k=0}^{N}\mathbb{E}\left[\binom{X_i^*}{k}\right]\varepsilon(N-k, \Delta^k p_h(x+1), W^{(i)}, X_i) + \sum_{i=1}^{n}\lambda_i\delta(N, p_h(x+1), W^{(i)}, X_i^*).\end{aligned} \tag{4.61}$$

Proof. We deduce by induction. The theorem holds when $N = 0$. Suppose that we have proved for $0, \cdots, N-1$ with $N \geq 1$. By (4.38),

$$\mathbb{E}[h(W)] = \mathcal{P}_{\lambda_W}(h) + \sum_{i=1}^{n} \lambda_i \Big(\mathbb{E}[p_h(W^{(i)} + X_i^* + 1)] - \mathbb{E}[p_h(W+1)] \Big).$$

Now, we shall rewrite the expectation $\mathbb{E}[p_h(W^{(i)} + X_i^* + 1)]$ as an N^{th}-order expansion on W. To this end, We take the difference expansion at $W^{(i)}$ and then apply (4.55) to get

$$\begin{aligned}
&\mathbb{E}[p_h(W^{(i)} + X_i^* + 1)] \\
&= \sum_{k=0}^{N} \mathbb{E}\left[\binom{X_i^*}{k}\right] \mathbb{E}[\Delta^k p_h(W^{(i)} + 1)] + \delta(N, p_h(x+1), W^{(i)}, X_i^*) && (4.62)\\
&= \sum_{k=0}^{N} \mathbb{E}\left[\binom{X_i^*}{k}\right] \Bigg(\mathbb{E}[\Delta^k p_h(W+1)] + \sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N-k}} \mathbb{E}[\Delta^{|\mathbf{J}|+k} p_h(W+1)] \prod_{l=1}^{d} \mathbb{E}\left[\binom{X_i}{j_l}\right] \\
&\quad + \varepsilon(N-k, \Delta^k p_h(x+1), W^{(i)}, X_i) \Bigg) + \delta\big(N, p_h(x+1), W^{(i)}, X_i^*\big). && (4.63)
\end{aligned}$$

To get the right order, (4.63) is obtained by making $(N-k)^{\text{th}}$ expansion of $\mathbb{E}[\Delta^k p_h(W^{(i)} + 1)]$ in (4.62). The first term in the bracket in (4.63) when $k = 0$ equals $\mathbb{E}[p_h(W+1)]$. The other summands in the first term when $k \geq 1$ can be regrouped with the second term by introducing the notation

$$\sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^0 \\ |\mathbf{J}|\leq N-k}} \mathbb{E}[\Delta^{|\mathbf{J}|+k} p_h(W+1)] = \mathbb{E}[\Delta^k p_h(W+1)].$$

Then (4.63) yields

$$\begin{aligned}
&\mathbb{E}[p_h(W^{(i)} + X_i^* + 1)] - \mathbb{E}[p_h(W+1)] \\
&= \sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}|} p_h(W+1)] \prod_{l=1}^{d} \mathbb{E}\left[\binom{X_i}{j_l}\right] && (4.64)\\
&+ \sum_{k=1}^{N} \mathbb{E}\left[\binom{X_i^*}{k}\right] \sum_{d\geq 0}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d \\ |\mathbf{J}|\leq N-k}} \mathbb{E}[\Delta^{|\mathbf{J}|+k} p_h(W+1)] \prod_{l=1}^{d} \mathbb{E}\left[\binom{X_i}{j_l}\right] && (4.65)\\
&+ \sum_{k=0}^{N} \mathbb{E}\left[\binom{X_i^*}{k}\right] \varepsilon(N-k, \Delta^k p_h(x+1), W^{(i)}, X_i) + \delta(N, p_h(x+1), W^{(i)}, X_i^*) && (4.66)
\end{aligned}$$

Taking the sum of (4.64) and (4.65), we get

$$
\begin{aligned}
&(4.64)+(4.65)\\
&=\sum_{d\geq 1}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}|}p_h(W+1)]\left(\prod_{l=1}^d \mathbb{E}\left[\binom{X_i}{j_l}\right]\right)\\
&+\sum_{d\geq 0}(-1)^d \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^{d+1}\\ |\mathbf{J}|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}|}p_h(W+1)]\left(\prod_{l=1}^d \mathbb{E}\left[\binom{X_i}{j_l}\right]\right)\mathbb{E}\left[\binom{X_i^*}{j_d}\right]\\
&=\sum_{d\geq 1}(-1)^{d-1} \sum_{\substack{\mathbf{J}=(j_l)\in\mathbb{N}_*^d\\ |\mathbf{J}|\leq N}} \mathbb{E}[\Delta^{|\mathbf{J}|}p_h(W+1)]\left(\prod_{l=1}^d \mathbb{E}\left[\binom{X_i^*}{j_l}-\binom{X_i}{j_l}\right]\right).
\end{aligned}
$$

By induction, we replace $\mathbb{E}[\Delta^{|\mathbf{J}|}p_h(W+1)]$ by its $(N-|\mathbf{J}|)^{\text{th}}$ order expansion $C(N-|\mathbf{J}|,\Delta^{|\mathbf{J}|}p_h(x+1))+e(N-|\mathbf{J}|,\Delta^{|\mathbf{J}|}p_h(x+1))$ to obtain (4.60) and (4.61). $\square$

Corollary 4.3.17 *We have the first two orders expansions*

$$C(1,h)=\mathcal{P}_{\lambda_W}(h)+\frac{\lambda_W}{2}\mathcal{P}_{\lambda_W}(\Delta^2 h)\mathbb{E}[X_I^*-X_I]$$

and

$$
\begin{aligned}
C(2,h)=C(1,h)&+\frac{\lambda_W^2}{8}\mathcal{P}_{\lambda_W}(\Delta^4 h)\big(\mathbb{E}[X_I^*]-\mathbb{E}[X_I]\big)^2\\
&+\frac{\lambda_W}{6}\mathcal{P}_{\lambda_W}(\Delta^3 h)\Big(\mathbb{E}[X_I^*(X_I^*-1)]-\mathbb{E}[X_I(X_I-1)]-2\mathbb{E}[X_I]\mathbb{E}[X_I^*-X_I]\Big).
\end{aligned}
$$

Remark 4.3.18 It is not difficult to verify that the Bernoulli case in the previous section is a special case of the above corollary.

Bibliography

[1] L. Andersen, J. Sidenius, and S. Basu, *All hedges in one basket*, Risk **16** (2003), 67–72.

[2] A. Antonov, S. Mechkov, and T. Misirpashaev, *Analytical techniques for synthetic CDOs and credit default risk measures*, NumeriX, 2005.

[3] A. D. Barbour, *Asymptotic expansions based on smooth functions in the central limit theorem*, Probability Theory and Related Fields **72** (1986), 289–303.

[4] ———, *Asymptotic expansions in the poisson limit theorem*, Annals of Probability **15** (1987), 748–766.

[5] A. D. Barbour, L. H. Y. Chen, and K. P. Choi, *Poisson approximation for unbounded functions. I. Independent summands*, Statistica Sinica **5** (1995), no. 2, 749–766.

[6] A. D. Barbour and P. Hall, *On the rate of Poisson convergence*, Mathematical Proceedings of the Cambridge Philosophical Society **95** (1984), no. 3, 473–480.

[7] A. D. Barbour, L. Holst, and S. Janson, *Poisson approximation*, Oxford University Press, 1992.

[8] A. Bélanger, S. E. Shreve, and D. Wong, *A general framework for pricing credit risk*, Math. Finance **14** (2004), no. 3, 317–350.

[9] T.R. Bielecki and M. Rutkowski, *Credit risk : modeling, valuation and hedging*, Springer-Verlag, 2002.

[10] C. Bluhm, L. Overbeck, and C. Wagner, *An introduction to credit risk modeling*, Chapman and Hall / CRC, 2003.

[11] I. S. Borisov and P. S. Ruzankin, *Poisson approximation for expectations of unbounded functions of independent random variables*, Annals of Probability **30** (2002), no. 4, 1657–1680.

[12] P. Brémaud and M. Yor, *Changes of filtrations and of probability measures*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **45** (1978), no. 4, 269–295.

[13] R. Bruyère, R. Cont, R. Copinot, L. Fery, C. Jaeck, and T. Spitz, *Credit derivatives and structured credit*, John Wiley and sons, Ltd., 2006.

[14] U. Çetin, R. Jarrow, P. Protter, and Y. Yıldırım, *Modeling credit risk with partial information*, Annals of Applied Probability **14** (2004), no. 3, 1167–1178.

[15] L. H. Y. Chen, *Poisson approximation for dependent trials*, Annals of Probability **3** (1975), 534–545.

[16] ———, *Stein's method: some perspectives with applications*, Probability towards 2000 (New York, 1995), Lecture Notes in Statist., vol. 128, Springer, New York, 1998, pp. 97–122.

[17] L. H. Y. Chen and Q.-M. Shao, *A non-uniform Berry-Esseen bound via Stein's method.*, Probability Theory and Related Fields **120** (2001), 236–254.

[18] ———, *Stein's method for normal approximation*, An Introduction to Stein's Method, Lecture Notes Series, IMS, National University of Singapore, vol. 4, Singapore University Press and World Scientific Publishing Co. Pte. Ltd., 2005, pp. 1–59.

[19] Louis H. Y. Chen and K. P. Choi, *Some asymptotic and large deviation results in Poisson approximation*, Ann. Probab. **20** (1992), no. 4, 1867–1876.

[20] P. Collin-Dufresne, R. Goldstein, and J. Helwege, *Is credit event risk priced? modelling contagion via the updating of beliefs*, (2003), Working Paper, Carnegie Mellon University.

[21] P. Deheuvels and D. Pfeifer, *On a relationship between Uspensky's theorem and Poisson approximations*, Annals of the Institute of Statistical Mathematics **40** (1988), no. 4, 671–681.

[22] C. Dellacherie and P.-A. Meyer, *Probabilités et potentiel*, Hermann, Paris, 1975, Chapitres I à IV, Édition entièrement refondue, Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV, Actualités Scientifiques et Industrielles, No. 1372.

[23] ———, *Probabilités et potentiel. Chapitres V à VIII*, revised ed., Actualités Scientifiques et Industrielles, vol. 1385, Hermann, Paris, 1980, Théorie des martingales.

[24] A. Dembo, J.-D. Deuschel, and D. Duffie, *Large portfolio losses*, Finance and Stochastics **8** (2004), no. 1, 3–16.

[25] F. Diener and M. Diener, *Asymptotics of price oscillations of a european Call option in a tree model*, Mathematical Finance **14** (2004), no. 2, 271–293.

[26] D. Duffie, *Defaultable term structure models with fractional recovery of par*, Working paper, Standford University, 1998.

[27] ———, *First-to-default valuation*, Working paper, Standford University, 1998.

[28] D. Duffie and D. Lando, *Term structure of credit spreads with incomplete accounting information*, Econometrica **69** (2001), 633–664.

[29] D. Duffie and K.J. Singleton, *Modeling term structures of defaultable bonds*, The review of financial studies **12** (1999), no. 4, 687–720.

[30] D Duffie and K.J. Singleton, *Credit risk: pricing, measurement and management*, Princeton University Press, 2003.

[31] R.J. Elliott, M. Jeanblanc, and M. Yor, *On models of default risk,*, Mathematical Finance **10** (2000), 179–195.

[32] W. Feller, *An introduction to probability and its applications, 2nd ed.*, Wiley, New York, 1971.

[33] R. Frey and A.J. McNeil, *Modelling dependent defaults*, Working paper, 2001.

[34] K. Giesecke, *Default and information*, Journal of Economic Dynamics and Control, **to appear**.

[35] P. Glasserman, *Tail approximations for portfolio credit risk*, Journal of Derivatives (2004), 24–42.

[36] P. Glasserman, W. Kang, and P. Shahabuddin, *Large deviations of multifactor portfolio credit risk*, Mathematical Finance **to appear**.

[37] E. Gobet, *Analysis of the zigzag convergence for barrier options with binomial tree*, Prépublication 536, UMR7599 CNRS-Université Paris 6, 1999.

[38] L. Goldstein, *Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing*, Journal of Applied Probability **42** (2005), no. 3, 661–683.

[39] L. Goldstein and G. Reinert, *Stein's method and the zero bias transformation with application to simple random sampling*, Annals of Applied Probability **7** (1997), 935–952.

[40] ———, *Zero biasing in one and higher dimensions, and applications*, Stein's method and applications, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., vol. 5, Singapore Univ. Press, Singapore, 2005, pp. 1–18.

[41] L. Goldstein and Y. Rinott, *Multivariate normal approximations by Stein's method and size bias couplings*, Journal of Applied Probability **33** (1996), no. 1, 1–17.

[42] F. Götze and C. Hipp, *Asymptotic expansions in the central limit theorem under moment conditions*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **42** (1978), no. 1, 67–87.

[43] J. Gregory and J.-P. Laurent, *I will survive*, Risk **16** (2003), no. 6.

[44] X. Guo, R. Jarrow, and C. Menn, *A generalized Lando's formula: A filtration expansion perspective*, Working Paper, Cornell University, 2006.

[45] X. Guo, R. Jarrow, and Y. Zeng, *Information reduction in credit risks*, Working Paper, Cornell University, 2005.

[46] X. Guo and Y. Zeng, *Intensity process and compensator: A new filtration expansion approach and the Jeulin-Yor formula*, Working Paper, Cornell University, 2006.

[47] D. Heath, R. Jarrow, and A. Morton, *Bond pricing and the term structure of interest rates*, Econometrica **60** (1992), 77–106.

[48] C. Hipp, *Edgeworth expansions for integrals of smooth functions*, Ann. Probability **5** (1977), no. 6, 1004–1011.

[49] J. Jacod, *Calcul stochastique et problèmes de martingales*, Lecture Notes in Mathematics, vol. 714, Springer, Berlin, 1979.

[50] R. Jarrow and S. Turnbull, *Pricing options on financial securities subject to default risk*, Journal of finance **50** (1995), 53–86.

[51] R. Jarrow and F. Yu, *Counterparty risk and the pricing of defaultable securities*, Journal of Finance **56** (2001), no. 5, 1765–1799.

[52] M. Jeanblanc, *Credit risk*, Lecture notes, 2006.

[53] M. Jeanblanc and Y. LeCam, *Intensity versus hazard process*, Working paper in preparation, Université d'Evry, 2006.

[54] M. Jeanblanc and M. Rutkowski, *Modelling of default risk: mathematical tools*, Working paper, 2000.

[55] ———, *Default risk and hazard process*, Mathematical finance—Bachelier Congress, 2000 (Paris), Springer Finance, Springer, Berlin, 2002, pp. 281–312.

[56] M. Jeanblanc and S. Valchev, *Partial information and hazard process*, International Journal of Theoretical and Applied Finance **8** (2005), no. 6, 807–838. MR MR2170231

[57] T. Jeulin and M. Yor, *Grossissement d'une filtration et semi-martingale: formules explicites*, Séminaire de Probabilités, XII, Lecture notes in mathematics, 649, Springer, Berlin, 1978, pp. 78–97.

[58] S. Kusuoka, *A remark on default risk models*, Advances in Mathematical Economics **1** (1999), 69–82.

[59] D. Lando, *On Cox processes and credit risky securities*, Derivatives Research **2** (1998), 99–120.

[60] F. Laudenbach, *Calcul différentiel et intégral*, Éditions de l'École Polytechnique, Mars 2005.

[61] D.X. Li, *On default correlation: a copula function approach*, Working paper, Credit metrics group, 2000.

[62] R. Mansuy and M. Yor, *Random times and enlargements of filtrations in a brownian setting*, Springer-Verlag, 2006.

[63] R.C. Merton, *On the pricing of corporate debt: the risk structure of interest rates*, Journal of Finance **29** (1974), 449–470.

[64] P. A. Meyer, *Un cours sur les intégrales stochastiques*, Séminaire de Probabilités, X (Seconde partie: Théorie des intégrales stochastiques, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975), Springer, Berlin, 1976, pp. 245–400. Lecture Notes in Math., Vol. 511. MR MR0501332 (58 #18721)

[65] R. Nelsen, *An introduction to copulas*, Springer-Verlag, New York, 1999.

[66] V. V. Petrov, *Sums of independent random variables*, Springer-Verlag, 1975.

[67] P. Protter, *Stochastic integration and differential equations*, Springer-Verlag, Berlin, 1990.

[68] M. Raic, *Normal approximations by Stein's method*, Proceedings of the seventh young statisticians meeting (Andrej Mrvar, ed.), 2003, pp. 71–97.

[69] Y. Rinott and V. Rotar, *On Edgeworth expansions for dependency-neighborhoods chain structures and Stein's method*, Probability Theory and Related Fields **126** (2003), no. 4, 528–570.

[70] E. Rogge and P.J. Schönbucher, *Modelling dynamic portfolio credit risk*, Working paper, 2003.

[71] V. Rotar, *Stein's method, Edgeworth's expansions and a formula of Barbour*, Stein's method and applications, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., vol. 5, Singapore Univ. Press, Singapore, 2005, pp. 59–84.

[72] P.J. Schönbucher, *Credit risk modelling and credit derivatives*, Ph.D. thesis, University of Bonn, 2000.

[73] ______, *Credit derivatives pricing models, models, pricing and implementation*, Wiley, 2003.

[74] P.J. Schönbucher and D. Schubert, *Copula-dependent default risk in intensity models*, Working paper, University of Bonn, 2001.

[75] D. Shelton, *Back to normal*, Working paper, Citigroup, 2004.

[76] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proc, Sixty Berkeley Symp. Math. Statist. Probab., Univ. California Press, Berkeley., 1972, pp. 583–602.

[77] ______, *Approximate computation of expectations*, IMS, Hayward, CA., 1986.

[78] ______, *A way of using auxiliary randomization*, Probability Theory (L.H.Y. Chen, K.P.Choi, K. Hu and J.-H. Lou, edits), De Gruyter, Berlin., 1992, pp. 159–180.

[79] M. Jeanblanc et M. Rutkowski T.R. Bielecki, *Hedging of basket credit derivatives in cds market.*, 2005.

[80] O. Vasicek, *Limiting loan loss probability distribution*, Moody's KMV, 1991.

[81] Ch. Yoeurp, *Décompositions des martingales locales et formules exponentielles*, Séminaire de Probabilités, X (Seconde partie: Théorie des intégrales stochastiques, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975), Springer, Berlin, 1976, pp. 432–480. Lecture Notes in Math., Vol. 511.

[82] C. Zhou, *A jump-diffusion approach to modelling credit risk and valuing defaultable securities.*, Federal reserve board, 1996.